Material Quality Issues in Si and SiGe Molecular Beam Epitaxy

by David Warren Smith

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

University of Warwick,
Department of Physics.

August 1993

SUMMARY

MBE growth of Si and SiGe allows a high degree of control over doping profiles and of strained alloy semiconductor growth. The structures that can be formed as a result have many potential applications for new solid state physics studies and for commercial device exploitation. However, the electrical material quality of MBE grown Si and SiGe is a function of many complex, interrelated processes and has not been extensively studied.

In this work, the objectives have been to study the electrical quality of as-grown Si as a function of growth temperature and to investigate the material requirements for high mobilities (at low temperatures) in SiGe channel 2D hole gases. The former has been achieved by room temperature measurements of the generation lifetime using low temperature oxides, and the latter by parallel transport measurements at temperatures of 3K to 20K.

For the first time, generation lifetimes, τ_g, have been measured by an MOS capacitance transient technique where all post growth processing temperatures have been below the growth temperature, T_S. These lifetimes are believed to reflect as-grown Si quality. Values of τ_g in Si, grown in a V80 MBE system, were found to be $\sim$200 ns for T_S in the range 600 to 800 C. This study was limited by transient effects (that are believed to result from oxide breakdown) and no T_S dependence could be confirmed. A further study of Si grown in a V90S system revealed a complex, but repeatable, dependence of τ_g on T_S, with a minimum of $\tau_g \approx 10$ ns at mid-range T_S ($\sim$600 C). At high and low T_S, values of $\tau_g \approx 2$ µs were measured, as compared to values of $\tau_g \approx 10$ µs in control substrates. Importantly, post growth annealing has indicated that the T_S dependence is a not a thermal effect, i.e. it is growth related.

It has been necessary to develop some techniques for device processing in order to preserve the integrity of MBE structures. Some of these techniques were applied to the processing of MBE grown pn diodes, and values of τ_g measured by current- and capacitance-voltage methods. These values vary greatly, ranging from 1 ns to 10 µs from epilayer to epilayer and also varied by up to an order of magnitude across an epilayer. Values of τ_g obtained from control substrates were also low and inconsistent; therefore, this is likely to be due to the alteration of material quality during processing. The lifetime - T_s dependence is reconcilable with the as-grown results, but provides no confirmation due to the processing induced effects. The only high temperature process used was a 1000 C, 15 second implantation activation anneal, which was assessed by the MOS technique, and shown not to alter material quality.

All previous studies on SiGe channel 2D hole gases had reported relatively low mobilities, which were thought to be limited by alloy scattering. The material quality studies on this system, described here, have resulted in a maximum low temperature mobility greater than twice that previously reported for these structures. At the start of this study, mobilities were found to be limited by Cu contamination in the SiGe channel. The Cu was observed to redistribute in the structure as a result of a growth interruption and higher mobilities were obtained. Post growth annealing studies have indicated that this is a growth related effect. Modifications made to the growth technology significantly reduced the Cu concentration and further increased the mobilities.

The 4K 2D hole gas mobility is shown to increase with increasing growth temperature up to 640 C, which has been associated with a reduction in interface charge scattering and, again, post growth annealing has indicated that this is associated with a growth related process. By further increasing growth temperature to 850 C, a maximum mobility of $\sim$9000 $cm^2V^{-1}s^{-1}$ was obtained. At higher growth temperatures, mobilities degraded. Experiments, in which the SiGe channel width and 2D carrier concentration were varied, have indicated that the mobility, at high growth temperatures, is limited by both interface charge and long range ripple at the upper SiGe/Si interface, which can be described in terms of interface roughness scattering.

CONTENTS

Figure Captions

Tables

Acknowledgements

This work was carried out within the Advanced Semiconductor Research group at Warwick University. Inevitably, the author has benefited from working with all members of the group at some time and in some way. Particular thanks go to Bob Barlow, Nevil Mattey, Robin Biswas and Richard Houghton. I also wish to thank John Emeleus and Andrew Plews for their assistance with many hours of low temperature Hall measurements over the last two years. Thanks are also due to my supervisors Evan Parker and Terry Whall for their continual efforts to provide funding for this work and for helpful comments on the writing of this thesis.

I would like to thank Steve Taylor and Suo-Fa Wang of Liverpool University for low temperature oxide growths and the use of some analysis equipment. I have also gained some benefit from the use of the SERC microfabrication facility at Edinburgh University and from discussions about this work with Alan Gundlach and Jo DeLima. Thanks to Eddie Jones of the Rutherford Appleton Laboratory for his very patient assistance with mask production.

I must acknowledge the contributions of my wife, Julie, without whom the writing of this thesis would have been much harder, and my son, Charles, without whom it would have been a good deal easier.

Declaration

The work presented in this thesis was carried out by the author, except where specifically acknowledged in the text or acknowledgements. It is presented in accordance with the guide-lines laid down in the regulations of the University of Warwick, see Document Phys/PG3, 1988.

Some of this work has been, or is in the process of being, published:

" The pn Diode as a Diagnostic Tool for Si MBE"
D.W Smith, R.F.Houghton, T.E.Whall and E.H.C.Parker.
Thin Solid Films, 1990, 184, p177

"Growth Studies on $Si_{0.8}Ge_{0.2}$ Channel Two Dimensional Hole Gases"
D.W.Smith, C.J.Emeleus, R.A.A.,Kubiak, E.H.C.Parker and T.E.Whall.
Applied Physics Letters, 1992, 61(12), p1453

"High Hole Mobilities in a p-type modulation doped $Si/Si_{0.83}Ge_{0.17}$ Heterostructure"
T.E.Whall, D.W.Smith, A.Plews, P.J.Phillips, R.A.A.Kubiak, E.H.C.Parker.
Semiconductor Science and Technology, 1993, 8(4), p615

"Generation Lifetimes in Low Temperature Si MBE"
D.W.Smith, S-F Wang, S.Taylor and E.H.C.Parker.
Solid State Electronics, Accepted for publication, 1993.

CHAPTER ONE

INTRODUCTION

1.1 INTRODUCTION

In 1947 Bardeen and Brattain first observed amplification from a semiconductor device - a point contact transistor. In the intervening years a massive research effort has ensued, studying the properties of group IV semiconductors and homojunction devices, particularly in the silicon matrix. As a result, Si is without doubt the element which has been most extensively researched, in its solid phase, both experimentally and theoretically.

Some of the fundamental features of electronic transport in semiconducting systems have only been revealed in the last ten years: a number remain obscure. The first of these, the Integer Quantum Hall Effect (IQHE), was observed by Von Klitzing et al, 1980, in the inversion layer of a Si MOSFET, although apparent in results from Kawaji two years earlier. However, the major subsequent discoveries, the Fractional Quantum Hall Effect (FQHE), ballistic resistance effects and one and zero dimensionality, were seen in GaAs/$Ga_xAl_{1-x}As$ structures. New device concepts have also been almost exclusively proven in similar III-V systems. Yet these remain of relatively insignificant technological importance and the processing aetiology is not yet competitive with that of Si. The use of III-V semiconductors is due, in part, to the relative immaturity of low temperature Si/SiGe growth techniques.

The net rate of progress of Si technology is shown in Fig.1.1. This rate is currently approaching limitations of miniaturisation not of a technological nature but due to physics: device latch-up, dopant diffusion, density fluctuations and eventually quantum scale effects will tend to saturate the curve as critical dimensions reduce below 0.1 μm. Solutions must be sought other than by scaling down laterally. SiGe alloy

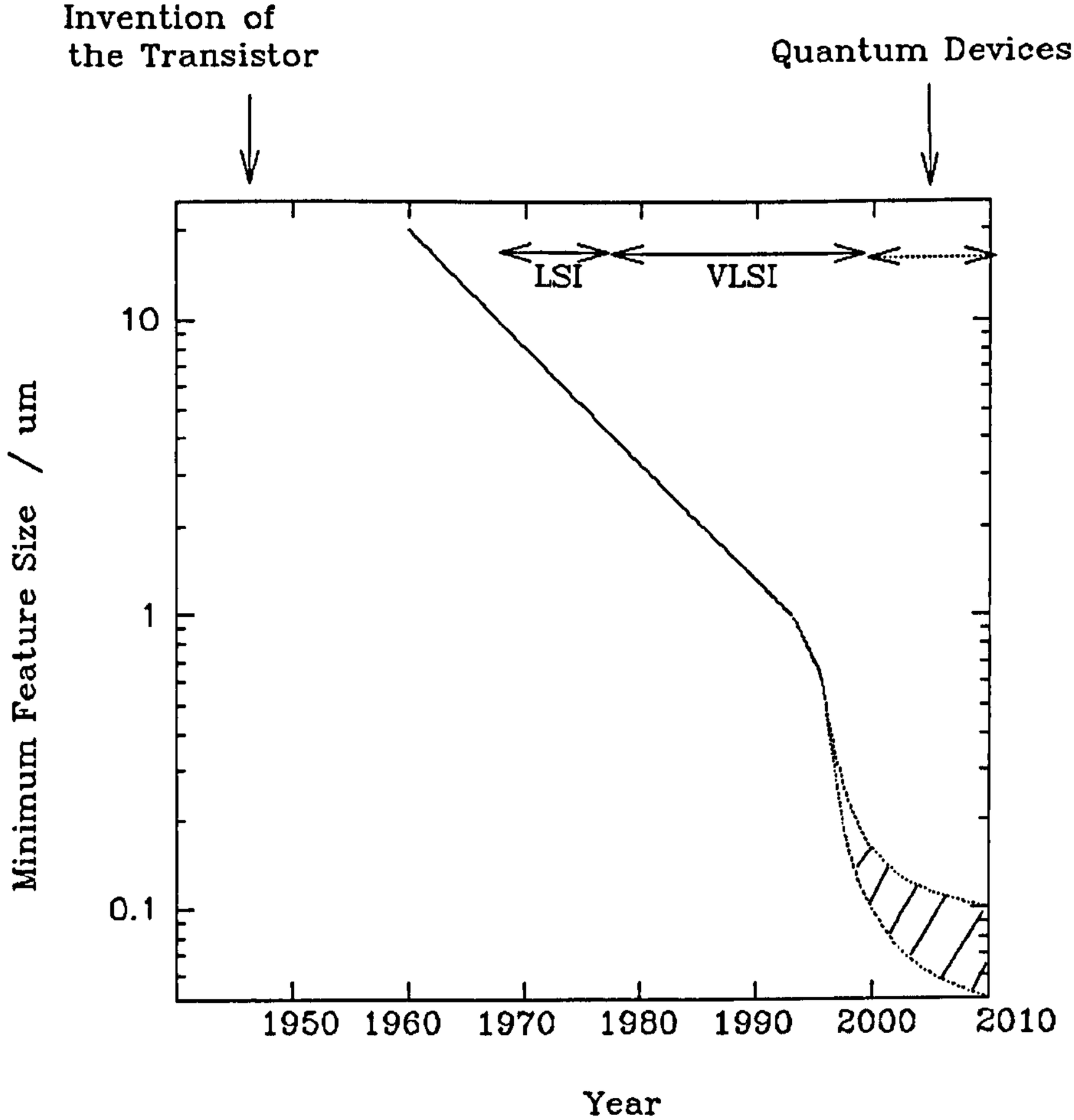


Fig. 1.1 Decrease of Si device minimum feature size since the invention of the transistor in 1947.

alloy systems offer one of the most exciting routes to extending the capability of current Si technology (e.g. the heterojunction bipolar transistor), including the possibility of a catalogue of devices based upon quantum effects (e.g. the SiGe quantum wire or the quantum dot), moving towards optically based devices (e.g. the Brillioun zone folded superlattice).

The semiconductor growth techniques required to produce 2D quantum systems use lower crystal growth temperatures, T_s, than those used for the production of bulk crystals. Growth at low T_s allows precise control of composition in the vertical axis and the retention of strain in alloy systems, but whilst this is a necessary condition for the exploitation of many novel device concepts, it is not sufficient. It is also necessary that the electrical material quality of the semiconductor is adequate for both device applications and for experimental low dimensional physics. Only limited information is available regarding the electrical quality of Si as a function of low growth temperatures. It is, however, known that crystalline quality is inferior to that grown at high T_s, i.e. by conventional techniques, and likely that metal gettering (a technique for material quality engineering discussed in Section 2.5) is hindered at low T_s. The issue of electrical material quality is addressed in this study with relation to Molecular Beam Epitaxy (MBE).

1.2 Objectives of This Study

Despite the potential of Si MBE there is little conclusive information regarding the formation of defects and the incorporation of unintentional impurities during growth. It is, however, known that some impurity incorporation and defect formation processes that occur during growth are not identical to well-studied solid state processes. There is even less information pertaining to the electrical quality of MBE grown material: this then is studied in this work. This lack of information is, in part,

due both to the difficulty of such studies and, in some part, economic reasons. MBE material quality is dependent on the complex interaction of three broad parameters: substrate preparation, growth schedules and growth technology. MBE has a low throughput, which makes systematic investigations of all of these parameter combinations extremely time consuming and costly. Two growth systems have been used during this work, initially a V80 system and later a V90S system. The substrate preparations relevant to each growth technology are reviewed in Chapter Two. The experimental nature of the MBE facility is such that it has only been practical to use growth schedules as the major variable parameter in this study, although some modifications of the V90S growth system were found necessary. Growth, or substrate, temperature, T_s, has been used as the intentional variable and has been found to be a very influential parameter for Si and SiGe material quality.

Accurate measurements of material quality are essential for this programme of work. Bulk-like majority carrier mobilities have been previously obtained at 77K and 300K in n- and p-type MBE grown Si (Kubiak et al, 1987), but these are relatively insensitive to deep levels in the band gap. Carrier lifetimes, however, are a very sensitive measure of material quality in Si. No systematic measurements of generation lifetimes as a function of Si MBE growth parameters have been reported and previous sporadic values have been obtained in MBE Si subject to high temperature post growth processing, which are likely to alter material quality. It has been necessary to modify an established technique in order to measure the lifetimes appropriate to as-grown Si. Using this technique, experiments have been undertaken to assess the 3D electrical properties of MBE grown Si as a function of the growth temperature. The measurement of carrier lifetimes, by various techniques, has been used to assess the effect of processing on as-grown MBE Si, which is an important aspect of the potential for MBE as a semiconductor growth technique.

The low temperature mobility of 2D charge gases are also highly sensitive to material quality and enhanced mobilities are of considerable interest, as will be discussed in Section 1.4. The material requirements for low scattering rates in 2D SiGe based structures have been investigated. This work contains one example of a novel structure in which greatly improved performance has been achieved as a result of material quality studies.

In this introduction, I shall outline the need for low temperature growth and the nature of the Si MBE process, particularly in terms of growth temperature. In order to illustrate the need for this study, it is also necessary to discuss the potential applications of Si and SiGe MBE for new devices and for new understanding in low dimensional physics. Examples of such applications are given with particular reference to the material quality issues. The thesis structure is given in the final section.

1.3 Si MBE GROWTH

Epitaxial growths of single crystal semiconductors that involve a transition from a vapour phase to the solid were pioneered in the 1960's. The proliferation of such techniques has led to a vast number of acronyms rivalled only by those for new devices, although, in common with these, the number of new concepts is considerably more limited. Hopefully this acronymania is curable.

Whilst MBE has proved a very useful tool for III-V semiconductors, considerably less research effort or progress has been in Si. The reasons for this are three-fold: many fundamental properties of Si are less convenient than those of say GaAs (i.e. lower carrier mobilities, indirect and larger band gap, higher intrinsic carrier concentrations, deeper impurity levels), the conventional technology of Si MBE (viz. solid source MBE) requires electron beam evaporation and, finally, if it is to be

viable, Si MBE must offer advantages over conventional Si processing. This latter requirement is both the strength and weakness of Si MBE as opposed to III V systems.

The technology of Si MBE has been well reviewed by Kaspar and Bean, 1988. Common to many low temperature growth techniques is the use of a very high vacuum and a heated substrate. The vacuum extends the mean free paths of the atoms to a few metres. These atoms, some of which may be ionised, impinge upon the substrate and may grow two-dimensionally as described by the Burton, Caberra and Frank theory, 1951. MBE is a non-equilibrium process in which absorbed atoms on the growth front can either desorb, or grow epitaxially by diffusion to a step and incorporation or by diffusion along a step before incorporation. Joyce 1988 has shown that the surface lifetime is sufficiently long for 10^6 jumps before desorption. This process is reviewed by Ghez and Iyer 1988.

The significant features of MBE are the very low deposition rates achievable ($<$ 1 nm/min) and the low growth temperatures (single crystal MBE growth of Si at 250 C has been reported by Jorke et al 1989). Low deposition rates enable very sharp dopant or alloy profiles to be obtained by co-evaporating dopants during Si growth. The epitome of such structures is the δ doped structure containing dopant spikes with FWHM of $<$ 1 nm. Low growth temperatures preserve the integrity of these profiles by minimising subsequent solid state diffusion. In addition to inhibiting diffusion, low T_s also reduces surface segregation and activated relaxation of strained alloy layers. If a heterojunction is formed by commensurate heteroepitaxy, the difference in band gaps must be accommodated at the interface. For some alloy systems, this provides a basis for the formation of a potential well and hence 2D electrical systems. Alloying allows semiconductor band gaps to be varied. (The band gap, E_g, of an alloy can be predicted to a first approximation by a simple linear interpolation of E_g for the constituent elements.)

The incorporation of unintentional impurities and crystalline defect formation will be topics much discussed in this work. For this reason it is necessary to outline the surface processes during growth to clarify the effect of T_s. For Si and Ge the incident flux, J_0, is much greater than the equilibrium flux re-evaporating from the surface, J_i, at all practical T_s, hence a condition of high supersaturation exists and there is a near unity condensation or sticking coefficient, S, given by,

$$S = \frac{J_0 - J_i}{J_0} \qquad ; S \rightarrow 1, \text{ for } J_0 >> J_i \qquad \textbf{1.1}$$

At 500 C, J_i is approximately 1 atom/cm^{2}s and at 900 C, J_i equals 10^{11} atoms/cm^{2}s with J_0 approximately 10^{15} atoms/cm^{2}s. Thus, the growth rate is independent of T_s, unlike Chemical Vapour Deposition (CVD). However, for sufficiently low T_s (less than 200 C), single crystal growth does not occur since the adatom mobility is small, so capture at monatomic steps is not efficient, resulting in 3D nucleation and polycystalline Si.

T_s has emerged as the critical parameter influencing the material quality of MBE Si and studies thus far have indicated many constraints on selecting an optimum T_s. According to Parker and Whall 1988, to reduce threading dislocation and s-pit (thought to arise from metal precipitation) densities, we require $T_s \approx 600$ C. According to Sidebotham et al 1988, for low concentrations of deep levels T_s must be greater than 700 C: for B doping, especially in narrow spikes, T_s must be less than 600 C: for Sb doping T_s should be greater than 700 C, preferably 800 C. High T_s is likely to enhance gettering processes. Clearly, more information regarding electrical quality as a function of T_s is required, as is a means of continually monitoring its role, particularly as the causes of some of the above considerations are mitigated or removed. The effect of subsequent processing on material quality will be a factor of equal importance to device quality. This must also be subject to routine appraisal. Methods for this appraisal are discussed in Chapter Four.

1.4 APPLICATIONS OF Si AND SiGe MBE

The incentives for assessing, and improving, material quality of any semiconductor system, are to improve device performance and to enable experimental studies of solid state physics. This section illustrates some of the advantages that low temperature growth techniques can provide in both of these areas and highlights the material quality issues.

1.4.1 Devices Requiring Low Growth Temperatures

Bulk semiconductor devices commonly in use can be assigned to three broad categories (excluding opto-electronic and Gunn effects). Angello 1958 classified them in terms of, essentially, the complexity of manufacture, i.e. the number of junctions, but a more useful approach for our purposes is to consider the mode of operation, which yields field effect devices, bipolar devices and barrier unipolar devices. The advent of low temperature growth methods have vastly expanded the number of devices in each category and also extended the number of fundamental categories. The devices discussed below all contain very narrow dopant profiles or strained alloy heterojunctions.

The recently proposed devices which fall under conventional descriptions nearly always mimic an existing device. They do so by moving the region of charge carrier control and transmission away from a surface interface to a homo or hetero junction within the semiconductor. Illustrative examples of this include the δ doped FET; here, the inversion layer at the $Si:SiO_2$ interface is replaced by a few monolayers of dopant, the δ layer, buried a few tens of nm below the surface. The free carrier concentration in this conducting channel is then modulated by a gate, either MOS or Schottky (Fig.1.2). This has, potentially, varied benefits: surface states are avoided,

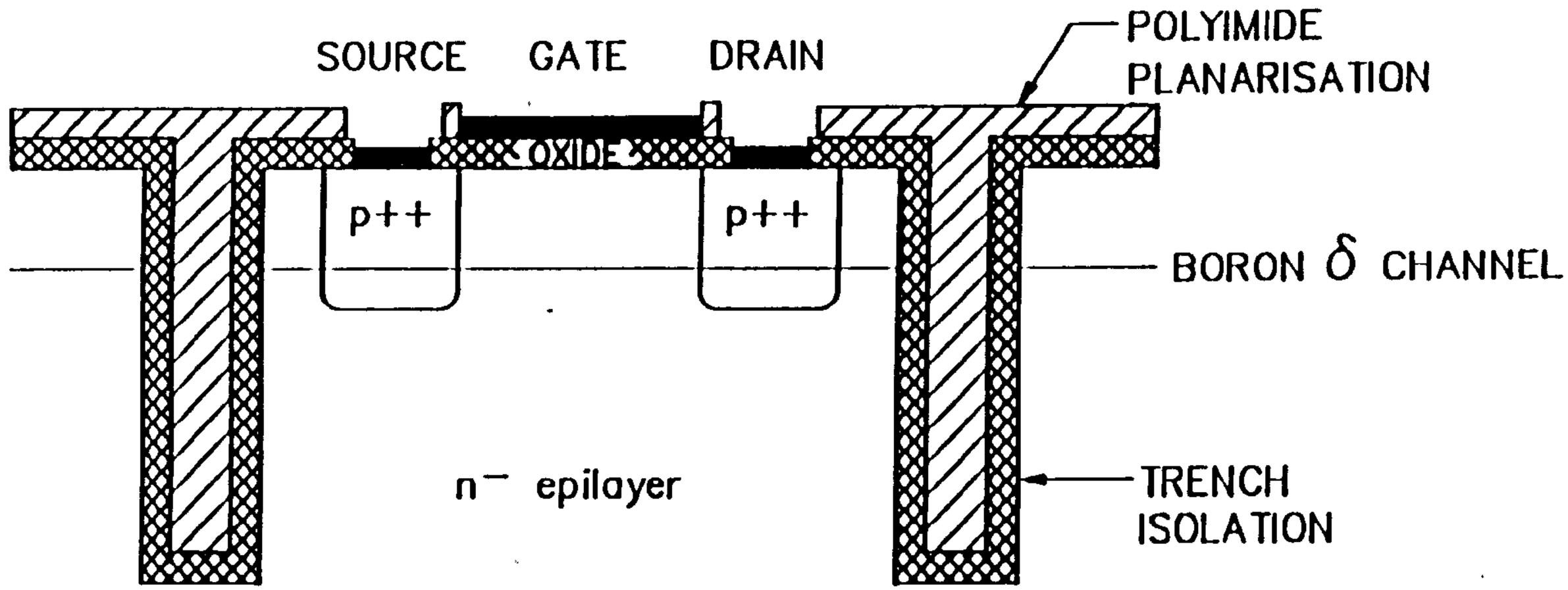


Fig 1.2 Cross sectional schematic of a B delta FET (processing and mask design by the author).

the gate capacitance is approximately constant, mobilities in the well sub bands may be higher than in bulk material and, of most importance, short channel (<0.1 μm) effects - electron injection into the oxide or density fluctuations - are reduced. The high electron mobility transistor (HEMT) is based upon similar principles, the confinement here provided at a heterojunction.

Bipolar devices have most promisingly been modified in the form of the Heterojunction Bipolar Transistor (HJBT), whereby the strained base is placed between two heterojunctions (Fig.1.3). This 'new' idea was included in Shockleys' original 1948 patent. Given favourable band offsets such as in Si:SiGe, the narrow band gap base allows the designer to use a thin base, but to dope it higher than the emitter to minimise the RC time constant without increasing the base-emitter current, which has produced f_t values currently approaching 100 GHz from SiGe base devices. Internal grading of the base alloy to enhance the built-in field will also result in shorter base transit times.

Barrier unipolar devices have been primarily restricted to the Schottky diode despite the very fast switching speeds resultant from no minority carrier storage effects. A class of devices that will be prove to be of importance is those based upon hot electron transport, which can be achieved by the use of thin dopant spikes in epitaxial structures. Their importance is twofold; firstly, because operating frequencies may be very high $> 10^{10}$ Hz (although the reasons for this may be subtle) and, secondly, because hot electron effects may be unavoidable in the next generation of devices. These devices are particularly interesting for Si MBE growth, since they are majority carrier devices and impurity scattering is diminished by virtue of the saturated electron velocities; material quality is, therefore, of less significance.

Proposed or realised devices that fall into new categories are the quantum wire FET (1D) and quantum dot (0D). However, the most thoroughly investigated is the resonant tunnelling diode. Since Sollner 1984 observed negative differential resistance

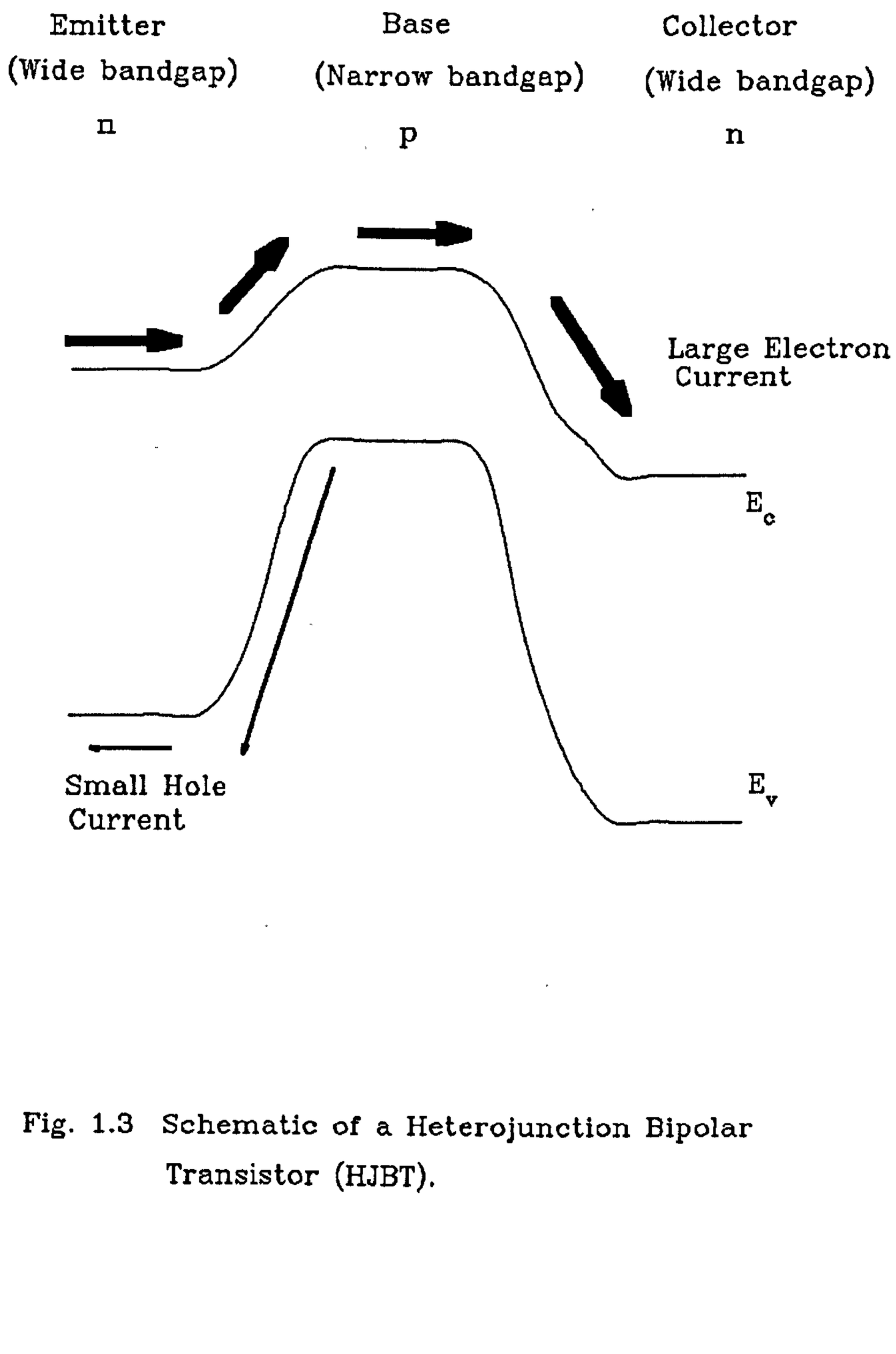


Fig. 1.3 Schematic of a Heterojunction Bipolar Transistor (HJBT).

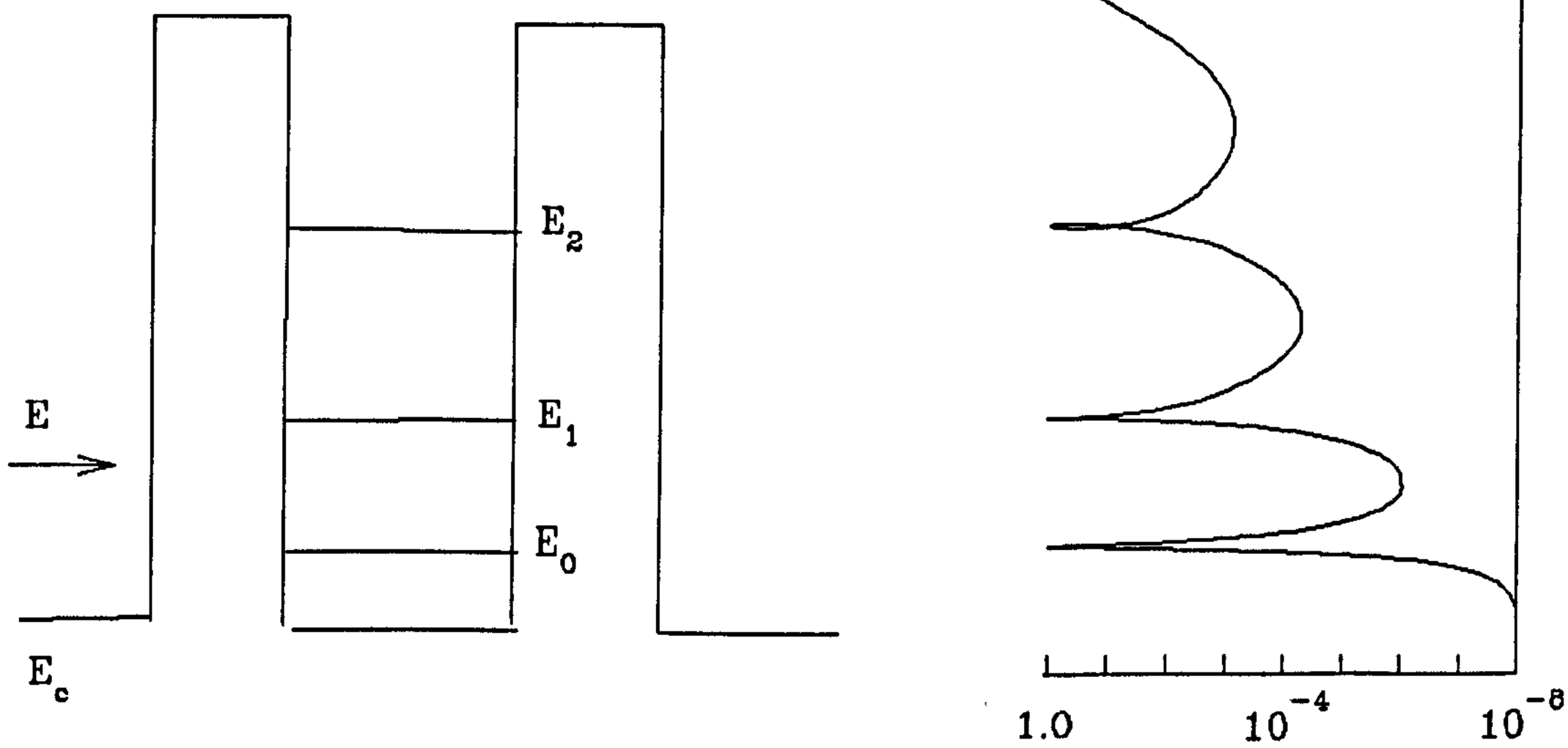
(NDR) from transverse transport measurements across a double barrier quantum well in the I-V curve, rather than, as was usual, in the first derivative, this device has provoked interest as a microwave oscillator. Its operation has been variously described in terms of a Fabry Perot resonator, hence the generic name, or as a sequential tunnelling effect. The Fabry Perot adherents took the view that when the incident carrier energy coincides with unoccupied states in the well, the electron wavefunction becomes coherent across the barrier. Alternatively, if this is viewed as sequential tunnelling (Fig.1.4), the transmission coefficient contains peaks at values of incident carrier energy that conserve lateral momentum. These two descriptions have recently been shown to be equivalent. It is important in the context of this chapter to note that if the barrier is not free from scattering centres, then lateral momentum will not be conserved during tunnelling.

1.4.2 Low Dimensional Phenomena

The electronic transport phenomena listed in the introductory Section are only observable in two or one dimensional systems at low temperatures. By confining the electron wavefunction in the z direction (2D) using a potential well, the allowable energy states are bound, discrete and form sub-bands above the conduction band edge, or below the valence band edge for holes. It is worth briefly considering these phenomena in order to highlight the role of disorder and to make clear the motivation for studying material quality via SiGe heterojunction 2D electron (hole) transport.

The IQHE describes the transverse and longitudinal resistivities of a 2D gas in the presence of a B field. Classically

$$\rho_{xy} = \frac{B}{eN(E)} \qquad \text{with N(E) the density of states.} \qquad \textbf{1.2}$$



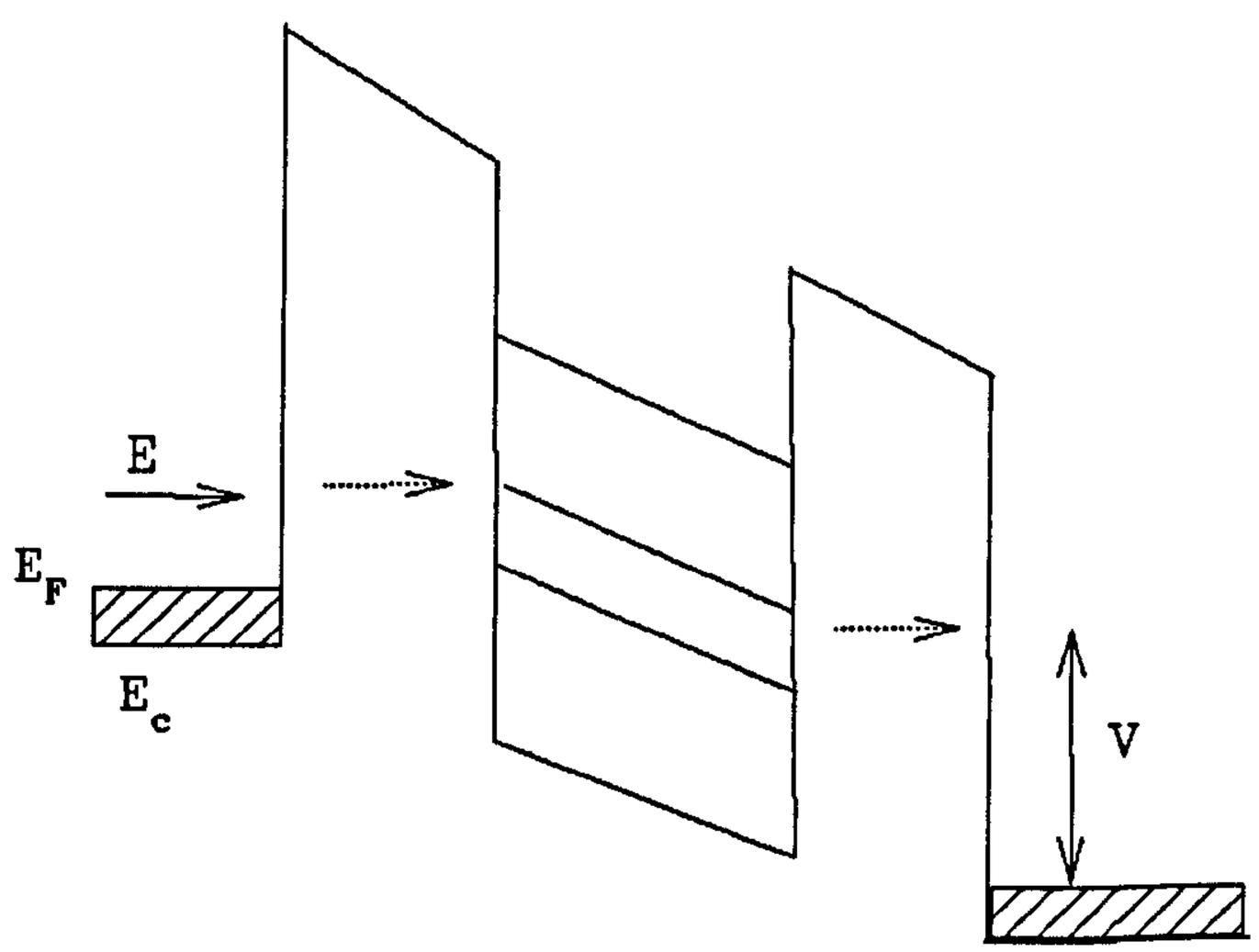


Fig 1.4 Band diagram for the resonant tunneling diode. Upper diagram is a schematic of the Fabry–Perot effect (barriers are transparent to electrons at resonant energies). Lower diagram illustrates sequential tunneling. (After Luyri, 1988)

However, experimentally, the dependence on B exhibits plateaux quantised with accuracy 1ppm at, $\rho_{xy} = \frac{h}{ie^2}$ where i = 1,2,3... .

At these values ρ_{xx} tends to zero ($< 5 \times 10^{-3}\ \Omega$ from Tsui 1982). The existence of singularities is explicable by considering the 2D density of states in a strong B field (Fig.1.5.a). As either the Fermi level is swept through the localised states, or, equivalently, as the states are swept by the B field through a fixed Fermi level, then plateaux occur. In order to explain the width of the plateaux, localisation must be invoked. Disorder in the system, which may be due to the random distribution of impurities, broadens the Landau levels, states in the tail of which are localised (Fig. 1.5 b). Laughlin 1981 has shown that the IQHE is independent of the strength of disorder. In the plateaux, σ_{xx} vanishes, since there are no vacant extended states within kT of E_f into which an electron can scatter, while σ_{xy} is determined by the density of occupied extended states. Various adequate analytic descriptions exist, derived via perturbation theory assuming a disorder potential, gauge invariance, percolation etc. Common to these is the assumption of non-interacting electrons. That this effect only of the representative new phenomena was first observed in a conventional Si device is not coincidental, since the effect is almost wholly independent of sample purity, provided that $\mu B > 1$. Indeed, the IQHE relies upon disorder, as such, it has become the basis for the international resistance standard - the Klitzing.

In very high mobility samples and for $T < 1K$, further structure is seen in the ρ_{xx} and ρ_{xy} data. When the B field is large enough, or the carrier density low enough, that the lowest Landau level is partially populated, ρ_{xy} exhibits plateaux at $\rho_{xy} = \frac{h}{\nu e^2}$ where ν is a rational fraction.

This is the FQHE. There are a number of theoretical approaches to this, all of which require consideration of electron - electron interactions. Currently foremost amongst

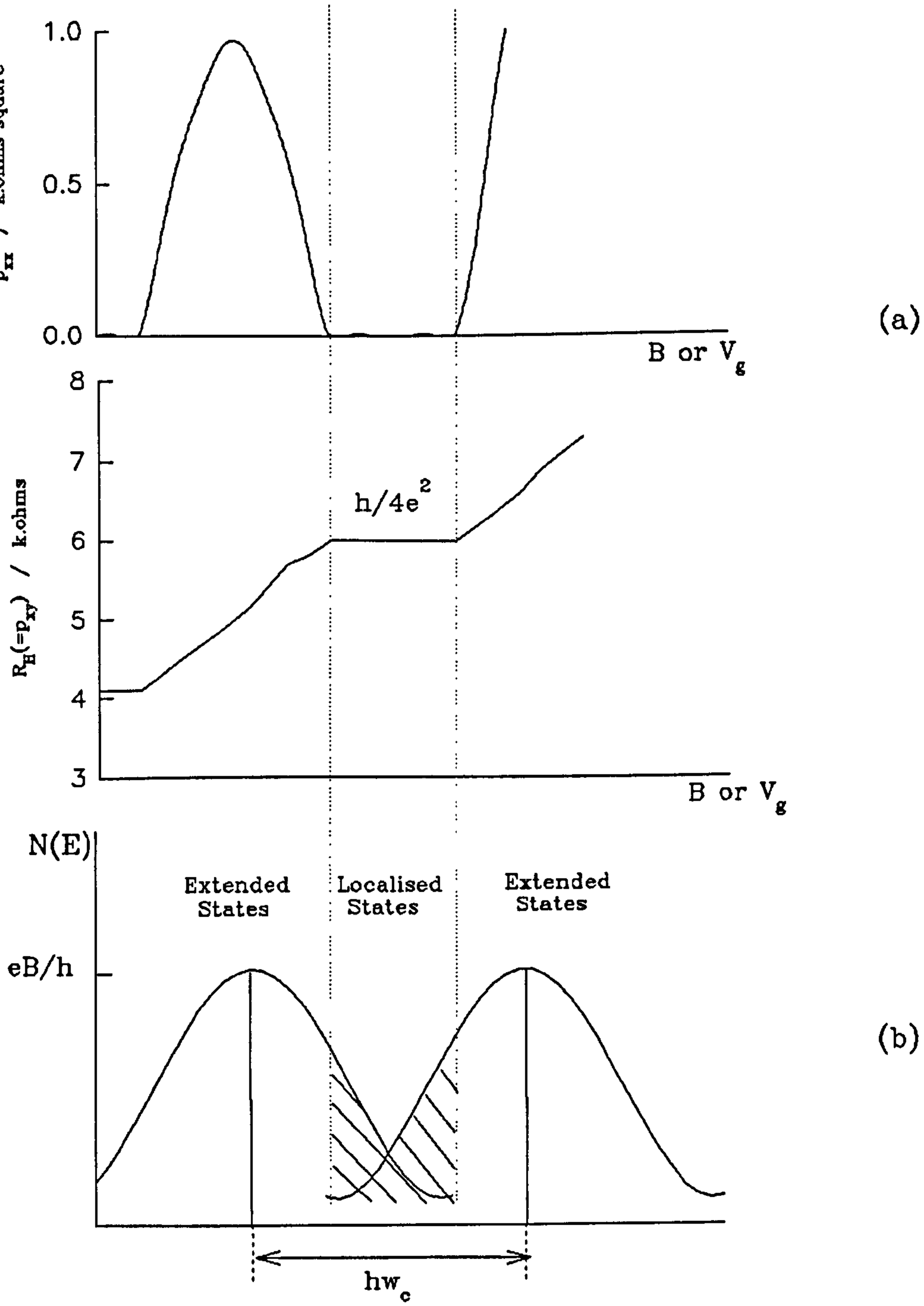


Fig. 1.5 Schematic illustration of the Integer Quantum Hall Effect. (a) illustrates the observed behaviour of the components of resistivity, (b) the density of states.

these descriptions is the formation of a Wigner electron liquid, whereby kinetic energy, confined by cyclotron resonance, is less than the electrostatic potential energy: this leads to the picture of quasi particles with fractional charge $q=e/\nu$. However, these descriptions are yet to be deemed conclusive: what is conclusive is that the accuracy of the FQHE is increased by lower scattering rates. Indeed, the effect is not experimentally observable in the presence of all but the minimum of disorder. As such the FQHE is only distantly related to the IQHE.

The ballistic resistance effect was initially demonstrated in a truly beautiful experiment jointly by Wharam et al 1988 and Van Wees et al 1988. A thin Schottky gate formed over a 2DEG was split (Fig.1.6 a), thereby producing a narrow channel for electron flow with width in the y direction controllable by applied gate bias. As the channel is narrowed, the resistance is seen to be quantised (Fig. 1.6 b), each quanta having value $h/2e^2$. This follows from consideration of the current contribution from one sub band

$I = \int n(v).e.v.dv$ with n(v) being the 1D density of states, v the velocity. **1.3**

We know that $n(v) = \frac{g_s m^*}{h}$, **1.4**

and by invoking conservation of energy, $m^*(v_i^2-v_f^2)=2eV$, so we have

$$I = \frac{2e^2V}{h}$$, assuming spin degeneracy of two. **1.5**

The total resistance of the channel is then given by

$$R = \frac{h}{2ie^2}$$, where i is the number of occupied sub bands. **1.6**

As a corollary, the placing of a second split gate separated from the first on the x axis by a distance s demonstrated that if s is $<$ 500 nm, the total series resistance was that of a single channel, whereas for s $>$ 500 nm, the resistances summed according to

Fig. 1.6 (a) Quantised resistance of a 1D channel.
(b) A schematic plan view of the split gate Schottky, where the metal arm width and gap are both ~0.5um.
(After Wharam et al 1988)

Kirchoffs law. This is perfectly comprehensible when considered in terms of an electron wave coherence length. Subsequently, this technique has directly demonstrated electron superposition via the Aharanov-Bohm effect etc. A range of semiconductor devices based upon electron wave behaviour offer many possibilities provided that large coherence lengths are achievable.

These phenomena form the basis of a range of proposed quantum devices, but require low charge carrier scattering rates. Clearly, if the controlled quantisation and related advantages that are realised by Si and SiGe MBE may be allied to existing Si processing, then a very powerful tool will be available. The maturity of Si processing will be a significant advantage for the production of low dimensional structures to probe further the physics of semiconductors. The production of novel high performance devices in a Si based technology may have considerable economic importance.

1.5 THESIS STRUCTURE

The previous discussions have, hopefully, made apparent that, without highly controllable band profiles and long range crystalline order with few unintentional impurities, the majority of low dimensional physical phenomena and useful novel device operations will not be observable. In order to meet these criteria, it is necessary to extend the carrier mean free path to greater than the physical dimension of the structure and to provide a high degree of control over carrier confinement. At the onset of this study, these two conditions were not jointly achievable using available Si MBE technology. This is the key to the relative status of III-V alloy structures and Si based structures and, therefore, to modern semiconductor physics. It is, however, now

just possible to achieve the aforementioned conditions using SiGe heterojunction structures grown by MBE, as will be discussed in this work.

It is necessary to review the constraints of Si MBE growth that determine material quality and also sample specifications. These matters are considered in Chapter Two. The post growth processing methods previously available equally limited the range of electrical measurements that could be taken. These methods, together with assessments of those developed and used by the author, are discussed in Chapter Three.

The project aims to lay down the groundwork for a study, of MBE Si quality from a device viewpoint. Hence the aim is to establish techniques for monitoring the 3D electrical quality of as grown MBE material, as a function of growth temperature, and the effect of low temperature processing steps. This should also establish some benchmark values of as-grown MBE material quality. This objective is best met by measurement of minority carrier lifetimes (recombination and generation) and the results of this study are discussed in Chapter Four.

The material quality requirements of the 2D SiGe system are considered via the remote doped p-type SiGe channel. Studies based upon low temperature parallel transport measurements, have been undertaken with the aim of producing high mobilities in the SiGe channel. These studies have necessitated modifications to both growth technology and growth schedules. This work is discussed in Chapter Five.

Finally, in Chapter Six, I summarise, conclude and consider future work.

CHAPTER TWO

MATERIAL ISSUES OF Si MBE GROWTH

2.1 INTRODUCTION

The design of semiconductor structures and their subsequent behaviour is fundamentally constrained or determined by the method of production. In this chapter, those constraints specific to Si MBE will be reviewed, particularly with regard to contamination: the consequences are primarily discussed in later chapters. The role of T_s, where information exists, is stressed. The post-growth device processing developed and the constraints resulting from this are discussed in the following chapter.

Device design requires a consideration of many factors of MBE growth as much as a consideration of the device physics. This chapter discusses doping - intentional and unintentional, substrate preparation, crystalline defects, metals and gettering as they are expected to affect device performance and growth specifications.

2.2 SUBSTRATE CLEANING

For 2-D growth to occur, the substrate surface must be 'atomically' clean. The primary aim of all cleaning procedures is to remove the native SiO_2. It has been shown that defect concentrations are critically related to surface cleanliness, certainly that inadequate SiO_2 removal yields highly defected material. Kimura and Lee (1975) carried out surface studies which suggest that high temperature in-situ sublimation cleans remove SiO_2, but leave, or introduce, significant C contamination which then acts as a nucleation site for 3-D growth. A strong correlation between residual C contamination and dislocations in the epilayer has been identified by McFee et al 1983,

but, surprisingly, no correlation between residual oxygen and dislocations was observed.

Various approaches to substrate cleaning have been used ranging from in-situ Ar sputtering (Bean et al 1977) to ex-situ chemical cleans. Methods of cleaning and the effects on defects are reviewed by Houghton et al 1987, although an equivalent study on the electrical consequences is not available. Cleaning procedures at Warwick in the early part of this study - in the V80 growth system - were to use a ex-situ 5 minute dip in 2.5% HF, and is currently - in the V90S - to use no ex-situ chemical cleans for 4" substrates and a full RCA clean for 3" substrates. An in-situ clean is essential for 2D growth, substrates in this study were subject to the Si cap procedure. Approximately 1 nm of amorphous Si is deposited at 350 C, T_s is increased to a 900 C for 5 mins with all cells shuttered, which results in oxide desorption via the reaction

$$Si + SiO_2 \rightarrow 2\ SiO\ (gaseous) \qquad 2.1$$

The pre-growth clean remains a possible parameter for a study of device quality.

In-situ desorption could only be carried out at 800 C in the V80 owing to limitations in the heater power supply. Irrespective of the ex-situ cleans used, dislocation concentrations remained $> 10^3$ cm^{-2}, whereas in the V90S, these were reduced to a few 10's per cm^2, which might suggest that residual SiO_2 dominates dislocation formation. However, this is not necessarily supported by some evidence in this study (see Section 3.6.4), regarding the difference between ex-situ cleans for 3" and 4" substrates.

2.3 CRYSTALLINE DEFECTS

Defect formation in single crystal growth by MBE as a function of T_s is complex. Crystallographic defects locally perturb an otherwise periodic lattice and, therefore, the Bloch functions which may then act as mid-gap centres. In addition, topographical defects ($>$ 1 μm diameter) are not tolerable for current VLSI systems. Minimising these defects is a further constraint upon growth conditions and hence device design.

There are three common defects in MBE Si as revealed by TEM, SEM and optical microscopy. To reveal these defects, these last two techniques require that these defects be revealed by a H_20:$Cr0_3$:HF (3:2:4) etch, (Schimmel 1979), the Si being preferentially oxidised by the $Cr0_3$ in regions of high strain or excess free electrons.

Dislocation densities in the V80 system were not reduced below 10^3 cm^{-2}. Having optimised cleaning procedures, dislocation density was apparently determined by T_s as shown in Fig. 2.1. Studies by the author on the material from the V90S system have shown dislocation densities approaching the detection limit for optical microscopy of 50 cm^{-2} and no T_s dependence has been observed (see Section 4.5.1.2). Whether this difference is due to a reduction in unintentional impurity incorporation, these then nucleating dislocations, or is due to the higher pre-clean substrate temperatures obtainable in the V90 S, or some other cause, is not known.

Saucer-Pits (S-Pits) have densities ranging from 10^2 cm^{-2} to 10^7 cm^{-2}: high densities of S-Pits can be seen prior to etching as the well known haze. The origin of these has been demonstrated by Pearce and McMahon 1977 to be caused by lattice relaxation about a precipitate. S-pits have been attributed solely to metal impurities, particularly Cu (Stacy et al 1981). A study by Parker and Whall, 1988, on material from the V80 system showed S-Pit densities in inverse proportion to the dislocation

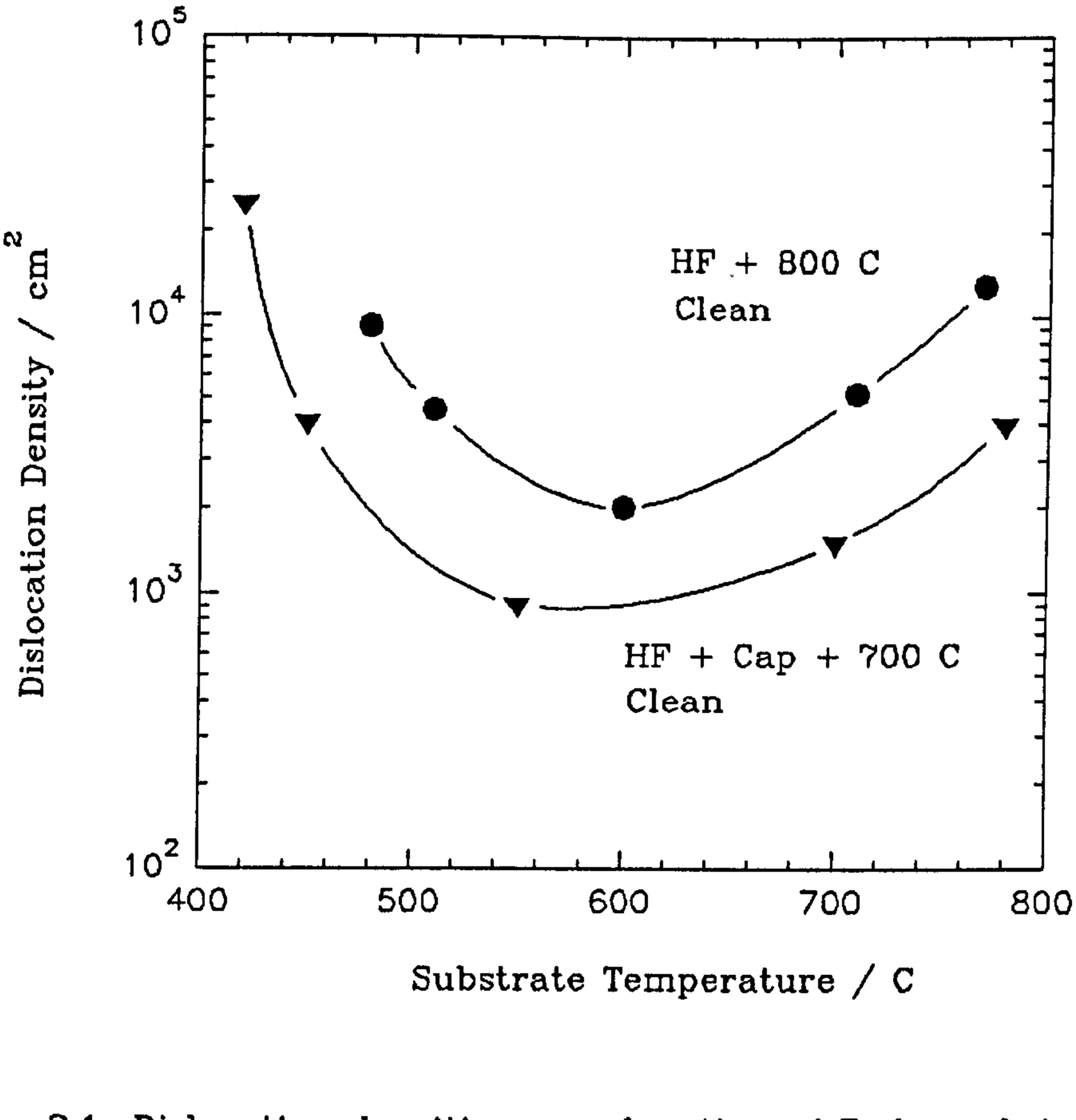


Fig. 2.1 Dislocation densities as a function of T_s for substrates prepared as indicated in the V80 system.
(After Parker and Whall, 1988)

density. This relationship has not been confirmed for a V90 S material where the author has observed no such correlation (see Section 3.4.6).

Particle induced defects (PID) have been much studied by Pindoria et al 1990. These are large, 0.5 μm diameter to 20 μm diameter, and are caused by Si particulates generated partly by the use of e-beam evaporation and are reduced by collimation. There is no observed dependence of PIDs on growth parameters other than growth times: the rate of arrival at the substrate is constant with time. As a result, PIDs do not affect device design, although, of course, they may well affect material quality by the introduction of crystalline disorder and local strain. Levels may be reduced to a few 10's cm^{-2} or may reach 10^6 cm^{-2} without collimation, but are typically of the order 100 cm^{-2}.

There is a dearth of information on the formation of the common crystalline defects in Si MBE; thus, we must rely upon defect etching and TEM to reveal defect levels but this, to date, has provided inconclusive evidence of the mechanisms causing the defects. Clues to these mechanisms were suggested by the T_s dependence found by Parker and Whall and referred to previously. However, as will be shown in Chapter Four, this dependence is not inherent to MBE growth, although it may be inherent to a particular growth system and clean schedule. Studies of defect creation are appropriate for the techniques of surface science: this study particularly focuses on the electrical activity of defects and this is discussed further in Chapter Four.

2.4 DOPING

2.4.1 Co-evaporation Doping

The complexity of impurity incorporation is seen via a consideration of intentional dopant behaviour. The choice of dopant in MBE is initially restricted by vapour pressure: if this is too low then high source cell temperatures are necessary, if too high at bakeout temperatures of 200 C (i.e. P) then it will act a source of background doping. Evaporation of dopant from the growth surface is characterised by the sticking coefficient. These coefficients are highly T_s dependent and can vary by four orders of magnitude from 500 to 900 C, which is in contrast to the case for the host species i.e. Si or Ge. Doping in this study is restricted to elemental boron and antimony.

Boron was thought to be very well behaved and to have a unity sticking coefficient dependent on the availability of incorporation sites, thus independent of T_s. Sb is a more problematic dopant due to its tendency to surface accumulation, thereby forming an adlayer at the growing surface, a process modelled by Pindoria et al 1990, suggesting that this behaviour is governed by the relative size of the dopant atom to that of the host lattice. The presence of an adlayer leads to both low doping levels and imprecise doping transitions, as doping continues from the surface reservoir during growth. This problem was solved jointly by Kubiak et al 1985 and Jorke and Kibbbel, 1985, via the application of a large negative potential (< 1 kV) to the substrate. This potential enhanced doping (PED) was originally attributed to secondary implantation from the adlayer by Si^+ ions. Houghton, 1991, has shown that the ion flux is not able to account for a 300% doping enhancement, whereas this is consistent with a model of Si^+ ions causing surfacial damage to the Si matrix, which increases the number of incorporation sites for Sb. The effectiveness of PED is illustrated in Fig 2.2. Two

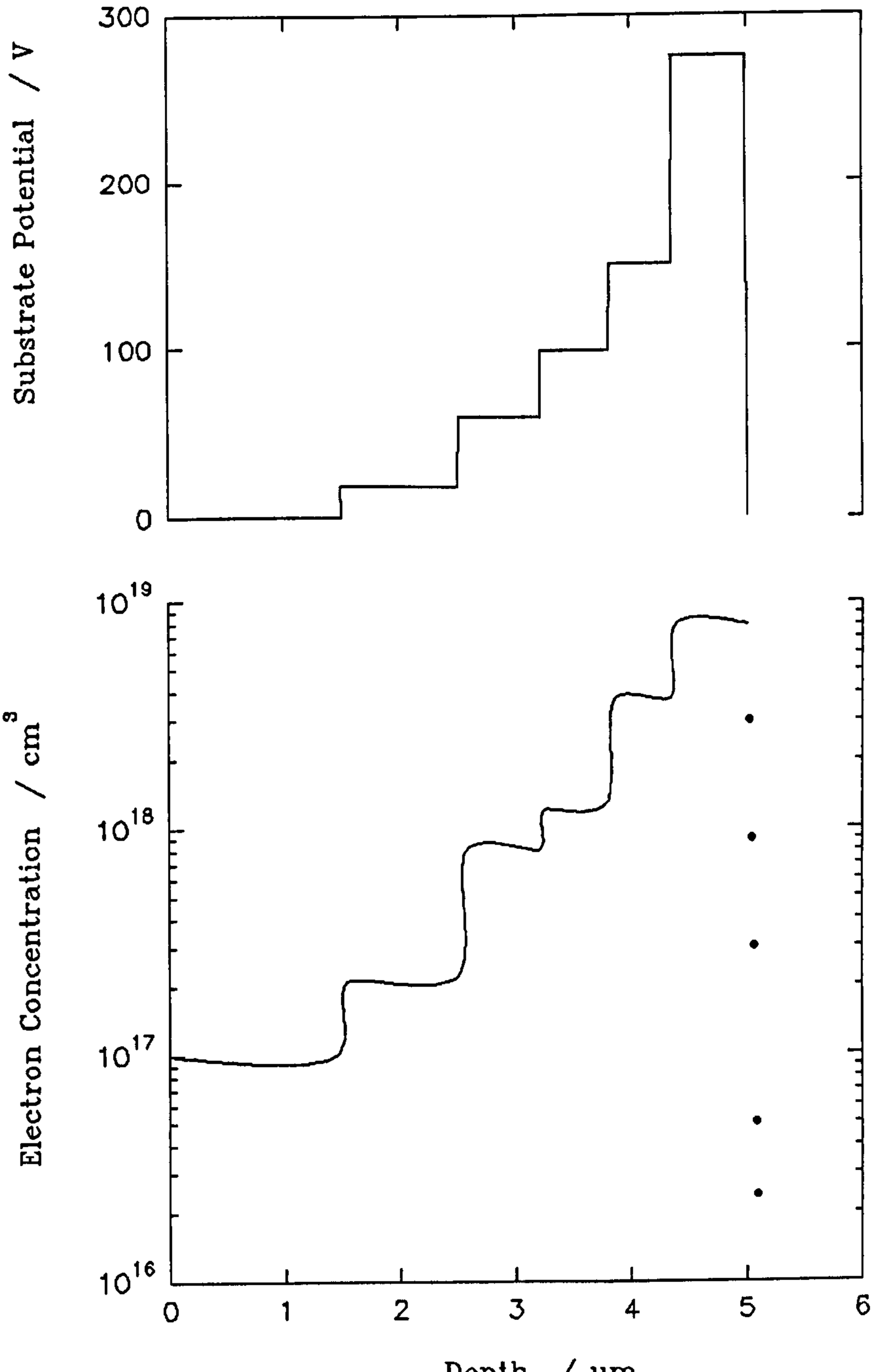


Fig. 2.2 Typical substrate voltage program and resulting Sb dopant profile, indicating the efficacy of PED (After Kubiak et al, 1985)

drawbacks remain regarding Sb; firstly, the T_s dependent sticking coefficient which presently precludes low T_s (< 600 C) and, secondly, the long thermal time constant of the source cell, which is of the order of hours.

An appreciation of the true difficulty in assessing impurity incorporation mechanisms comes from the p-type dopant B, which was thought to have a unity sticking coefficient. An early attempt at growth by MBE, reported by Smith et al 1988, of a bulk unipolar, or camel diode, (proposed by Shannon 1979), yielded poor device performance. This was attributed to B precipitation in a narrow B spike integral to the device, Fig 2.3. This observation led to a study of B solid solubility limits in Si as a function of growth temperature by Parry et al, 1992, from which it was found that the solid solubility limit decreases from 900 to 600 C in agreement with conventional dopant studies , Fig 2.4. Below 600 C the *apparent* solid solubility limit increases and at 500 C is higher than at 1000 C. Further studies by Parry and subsequently surface studies suggest that B forms an adlayer in a 2 x 2 surface reconstruction. The incorporation of this layer is governed by two mechanisms - at high T_s it is limited by thermal equilibrium and at low T_s is kinetically limited. This incorporation on cessation of B doping results in smearing of the dopant profile characterised by an accumulation length which is a strong function of T_s and growth rate, Fig 2.5 It is important to note that the high electrically active concentration obtained at low T_s is stable to post growth annealing of up to two hours at 900 C.

Finally, for completeness, brief mention is made of solid phase epitaxy (SPE), the technique used to produce delta layers. Si growth is interrupted, T_s lowered from 700 C to 400 C and a dopant flux is exposed to the surface until approximately monolayer coverage is achieved. A Si cap is then grown at 700 C for 2 mins which activates the δ layer. This achieves delta spikes of FWHM of 1 - 2 nm. The electrical and structural quality of these layers has been studied by Zeindl et al 1987 for Sb and Mattey et al 1990 for B.

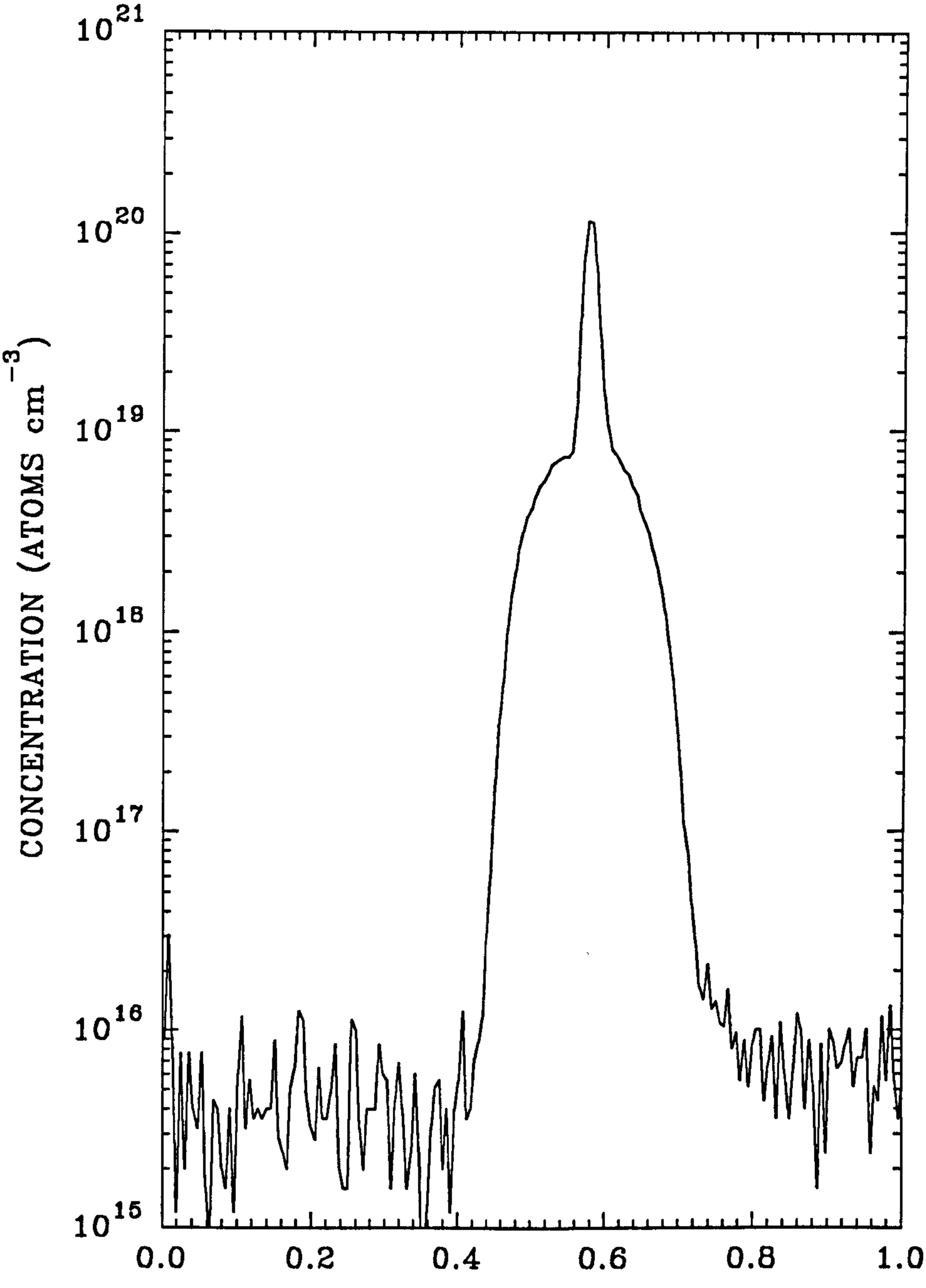


Fig. 2.3 SIMS profile of a nominally thin B doped layer grown by MBE, showing a precipitation spike above the solid solubility concentration.

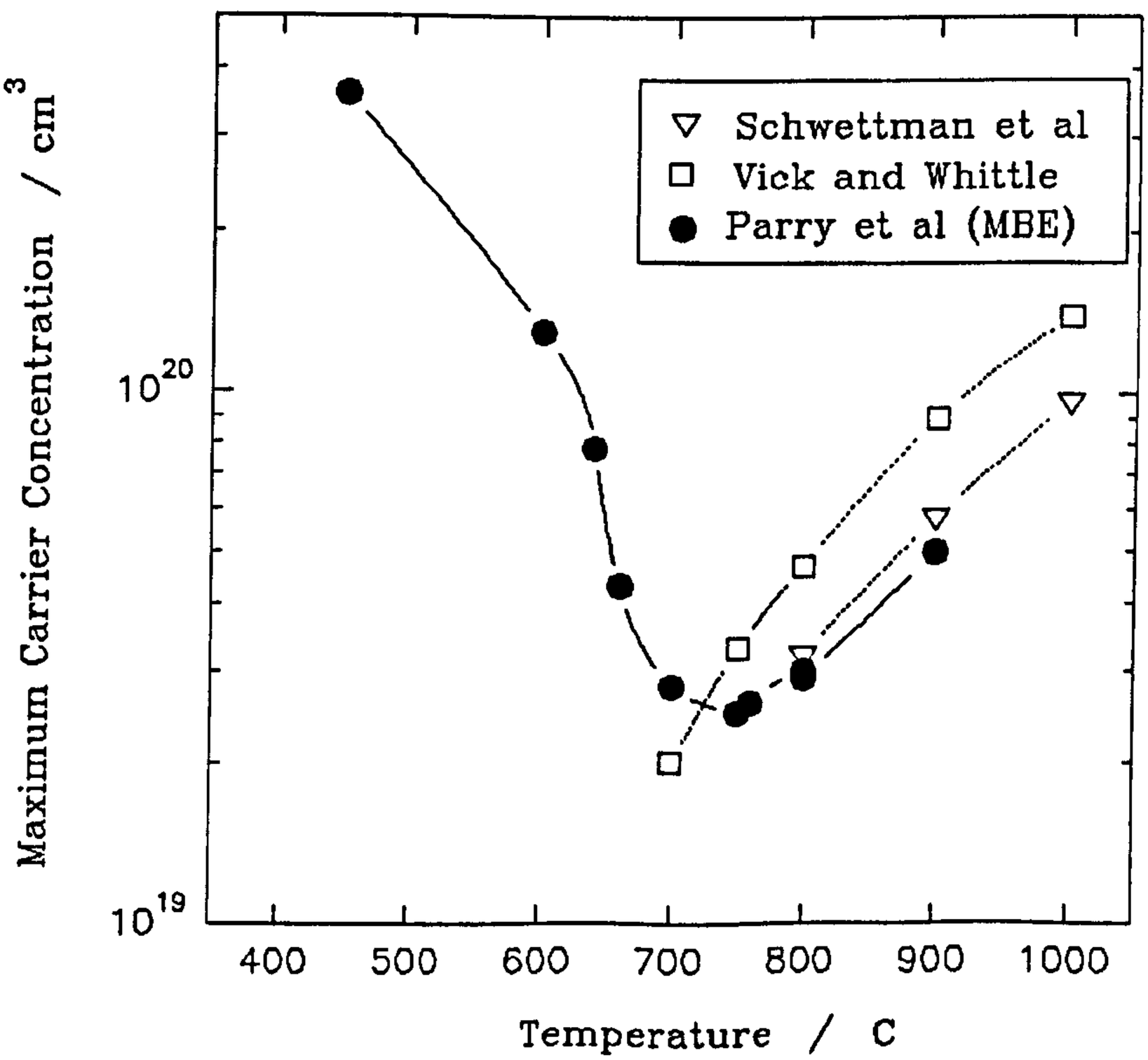


Fig. 2.4 Solid solubility limits of B in Si. Solid circles represent values obtained from MBE growth. (After Parry et al, 1992)

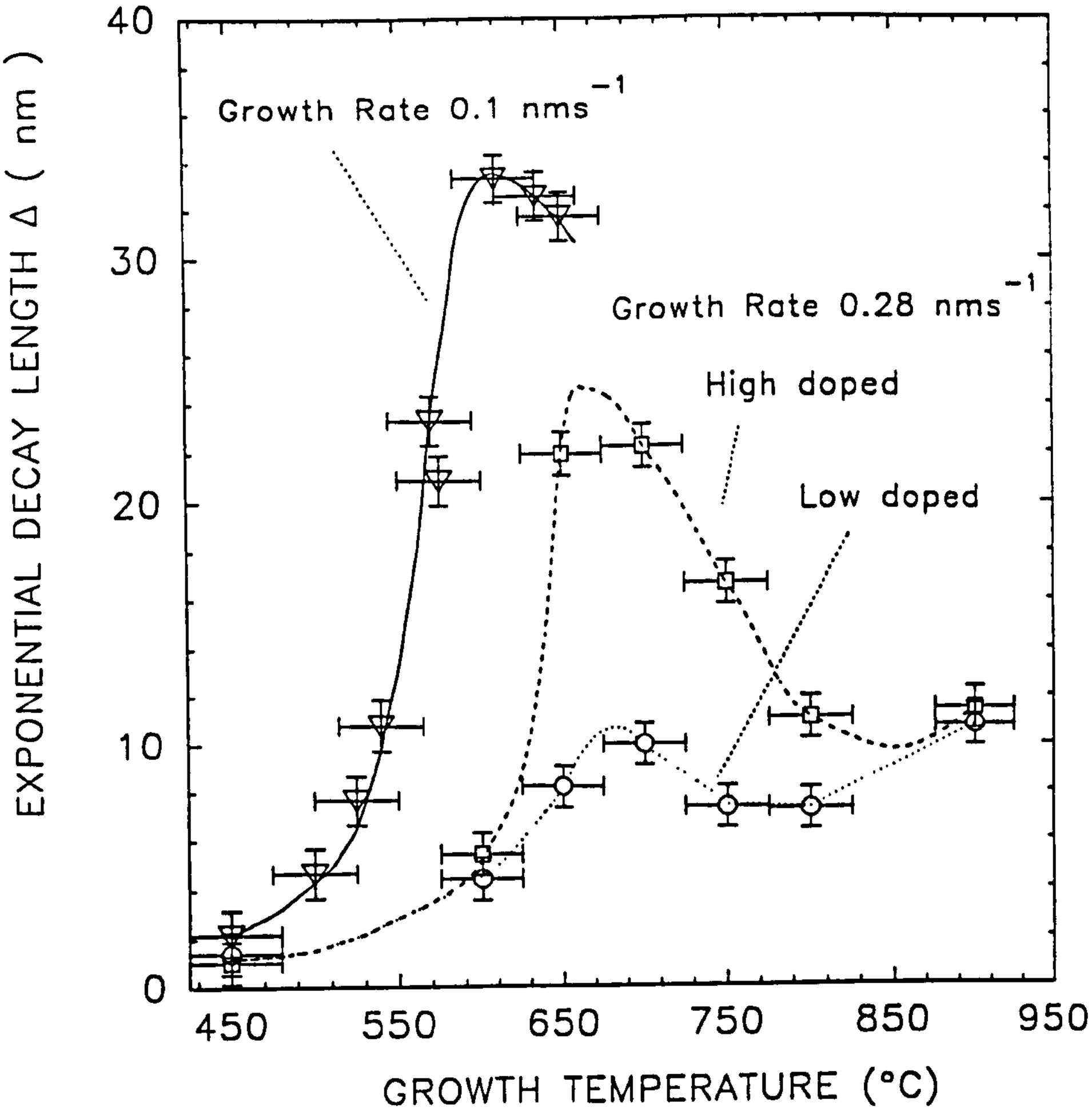


Fig. 2.5 The 1/e accumulation length of B in Si doped by co-evaporation. At low Ts the accumulation length is kinetically limited, at mid range T_s it is equilbrium limited and at high T_s it is diffusion limited. (After Parry et al 1991)

2.4.2 Unintentional Doping

Unintentional doping is a problem in MBE Si and plays a major role in this study. The B incorporation work discussed earlier illustrates the difficulty of modelling or predicting unintentional impurity behaviour in MBE grown Si. These difficulties are compounded by the likelihood that unintentional impurity concentrations will usually be below the detection limit of analytical techniques, that impurities may interact with crystalline defects and the possible formation of complexes. C and O are known to be present at levels up to 10^{18} cm^{-3}, and can increase by an order of magnitude at growth interrupts, the uptake increasing with T_s (Houghton, 1991). Much information regarding the electrical effect of these contaminants relates to the precipitate form. Higgs et al 1989 used PL on MBE Si and found free exciton luminescence from C, N and O related defects. These all annealed out fully at temperatures greater than 700 C for a 30 minute anneal, or at 900 C for a 2 minute anneal, and were not present in material with T_s greater than 800 C.

Background doping varies as an undetermined function of substrate and system cleanliness but ranges from 10^{14} to 10^{16} cm^{-3} and is n-type. Activation energies as derived from Hall temperature data led Houghton, 1991, to conclude that the residual dopant species in the V80 is P, presumably outgassing from stainless steel. This background compensation can have a considerable impact on sample design. It is important to note that background doping concentration has been found to be independent of T_s

Of even greater impact and interest is the B spike found at the epilayer-substrate interface (Fig 2.6). This has, typically, peak concentration of 10^{18} cm^{-3} with FWHM of 100-200nm. Various explanations have been proposed as to the source of this contamination. The level has been reduced by replacing borosilicate windows with quartz, but order of magnitude reductions are achieved by optimum cleaning

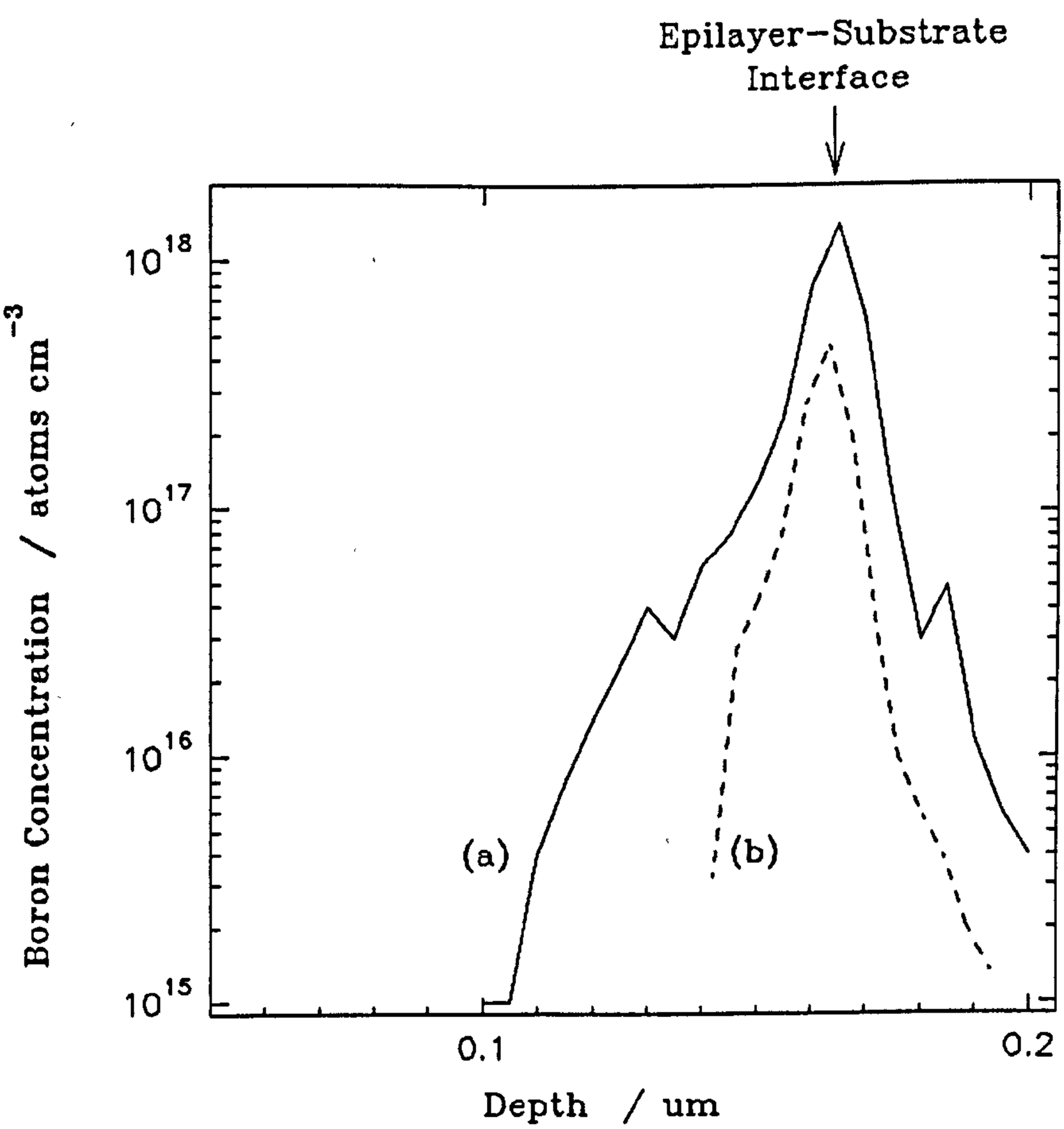


Fig. 2.6 SIMS B profile of two undoped epilayers, showing the B spike at the interface. (a) was grown with borosilicate glass viewports, (b) was grown with quartz viewports. (After Houghton, 1991)

procedures. It is believed that a B complex exists on as-received substrates, which disassociates to become electrically active for $T_s > 700$ C. Even when not electrically active the B complexes can affect electrical quality. This will be discussed later.

For Si MBE to prove practical, this interface represents a great restriction, second only to the high vacuums required and the restricted throughput . Its effects maybe mitigated by use of a buffer for parallel transport studies, but it remains a major distraction: for perpendicular transport, it may often be an overwhelming consideration for device design. The spike may have useful benefits (to be outlined in Section 2.5), but these are outweighed by the limitations it imposes.

2.4.3 Metals

The role of metals on carrier transport in Si is much studied. Other than in devices requiring large absorption of excess carriers i.e. optical detectors, metals have a wholly deleterious effect on device performance, which is why metallic levels in bulk material have been minimised to less than one part in 10^{10} for most metals. MBE Si has much higher metallic levels as is almost inevitable, given the necessary construction of metal vacuum systems enclosing heat sources. This is not the death knoll of MBE, since conventional processing techniques can also introduce high metal levels, which can be mitigated by gettering methods.

The metals primarily expected in MBE Si are those that are directly heated, specifically Cu from the Si source hearth, Ta from substrate heaters (now replaced by graphite meanders), both of which are shielded by quartz and W from the e-beam filament. The B source cell used has been shown to be a source of metallic contamination by the use of EDAX (electron dispersive x-ray analysis); the Ta ribbon contacts diffuse through the graphite. These have been replaced by another refractory metal, Mo, which diffuses at a lesser rate. This may not however be entirely

advantageous since Mo is a mid-gap trap with a highly deleterious effect on charge carrier lifetimes. In addition, the constituents of the stainless steel chamber (grade 316 stainless steel, Fe:Cr:Ni, 7:2:1) that are indirectly heated will outgas. These contaminants can be reduced by the use of water cooling, liquid N_2 cryopanels and regular bakeouts limited to 200 C by Viton seals, but there is evidence that they are not yet at acceptable levels.

Despite the high adatom mobility, to provide a high degree of matrix uniformity ($< 0.5\%$ variation) over a substrate, it is necessary to rotate the substrate, typically at π rad/min. This may have considerable ramifications for material quality due to the metal contamination introduced. Higashi et al , 1990, report a three-fold increase in minority carrier generation lifetime by stopping rotation, which they attribute to the reduction of metals generated by rotation (for a critical review of this work, see Chapter 4). Note that rotation has been shown to introduce C species by Farley et al 1987. Unfortunately, rotation is currently necessary unless uniformity is unimportant.

Two studies have been published using neutron activation analysis (NAA) - NAA is reviewed by Domenici et al 1986 - to establish metals species, and the level of contamination, in a V80 (another study by Liu, in a V80 growth system, is reported briefly by Marsh 1987). The first by Houghton et al, 1987, applies to the V80 system used in this study, the second by Van Gorkum et al, 1990, was also on a V80, both of which growth systems were extensively modified. Metallic levels found in the Houghton studies are shown in Table 2.1. The most interesting features are no clear dependence on T_s, with the possible exception of W, and the large (four order of magnitude) increase of metallics in a substrate loaded and heated but with no growth. Similar levels are found by Van Gorkum, but with a fall in Ta of one order of magnitude as monitored over one year of system use. Note that all of the metallic concentrations are *above* the minimum threshold concentration for the degradation of solar cells (Hopkins R 1986).

	Cu	Fe	Cr	Ta	Mn	W	Au	Na
Lower Limit / cm^{-3}	$1x10^{15}$	$1x10^{15}$	$3x10^{14}$	$2x10^{15}$	$3x10^{14}$	$7x10^{13}$	$1x10^{13}$	$6x10^{15}$
Upper Limit / cm^{-3}	$1x10^{16}$	$7x10^{15}$	$2x10^{15}$	$1x10^{16}$	$7x10^{15}$	$1x10^{15}$	$4x10^{13}$	$2x10^{16}$

Table 2.1 Minimum and maximum metal concentrations, in Si epilayers grown in the Warwick V80, as measured by NAA and AA. (For details of the growth schedules employed see Pindoria, 1990.)

Possible ways of removing the sources of some metals in the growth system are discussed in Chapter Five. The mechanisms by which metals are incorporated in the growth surface are beyond the scope of this study. However, the electrical effect of this metallic contamination is of relevance. A previous study by Sidebotham et al has used deep level transient spectroscopy (DLTS), Lang, 1974, to observe the mid-gap state concentration as a function of T_s , Fig 2.7. There is a two-order fall in total trap concentration as T_s is increased from 500 to 800 C. It must be stressed that this trend appears independent of defect density and of the total metallic concentration as determined by the aforementioned studies. This seems rather improbable. One possibility is that the primary source of deep levels is the heater. However, Pindoria has replaced the Ta ribbon heater with a graphite equivalent and observed deep level concentrations that fit the curve of Fig 2.7. In addition, this proposal would not explain the apparent independence of Ta concentration on T_s from the NAA study, and would imply a mechanism of incorporation that would have to render the other metal impurities electrically inactive, but not Ta. It could be argued that a maximum of 0.1% of the total metal concentration in MBE Si is electrically active at 500 C and 0.001% at 800 C. Such maxima assume no contribution from C, O, N or crystallographic defects and are not useful. It is clear that the very high metal concentrations do not have the full deleterious effect that could be expected. The incorporation of metals during growth must differ from that of diffusion into bulk materials, at least the majority being electrically inactive on the basis of these studies. Ravi, 1981, has suggested that dislocations do not act as mid-gap states unless decorated by a metal. If this were the sole cause of deep levels for MBE Si, then, given the relative densities of metals and deep levels, the trap concentration should follow the trend of dislocation density. From the preceeding, it is apparent that the situation must be more complex.

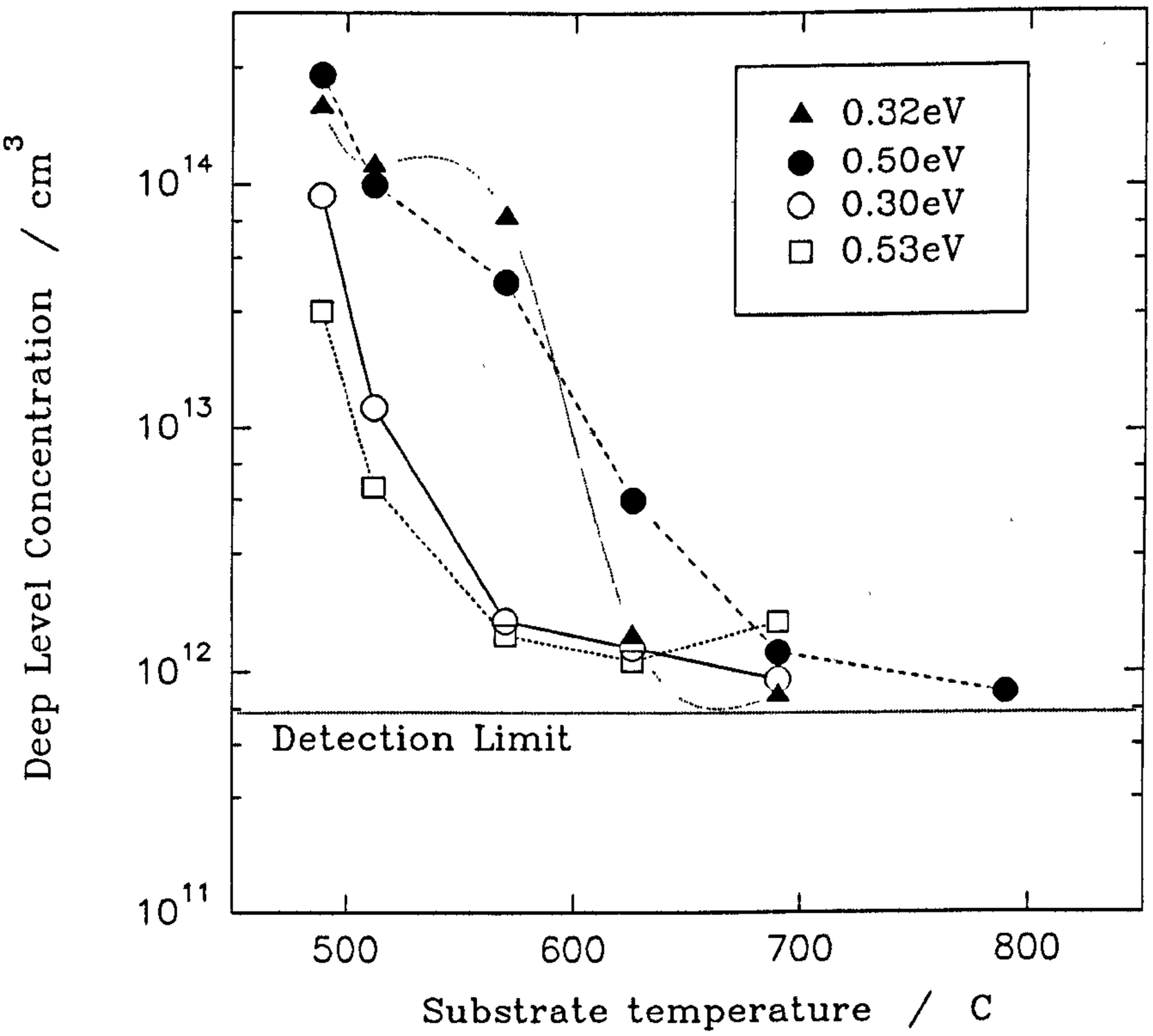


Fig 2.7 Deeep level concentration measured
by DLTS on V80 system Si
(After Sidebotham et al 1988)

DLTS is a finger print technique: from the capture cross sections, energy level and filling behaviour, it is possible to identify parameters characteristic of a given cause. Unfortunately, the vast number of possible contaminants, complexes interactions with defects and point defects themselves that contribute to the deep level library make positive identification difficult, particularly as MBE may produce previously unknown traps. Additionally, these parameters extracted from DLTS, whether from current experiment or in the 'library', are critically dependent on accurate and precise sample temperature measurement, and the necessary data manipulation is open to operator interpretation.

It is necessary that more electrical information is gathered, by far the most profitable approach being to monitor electrical material quality as functions of both time and growth chamber modification. Further, the effect of processing needs to be examined, some of which steps are not compatible with the need for Schottky contacts in DLTS i.e. heavy implantation doses. It was thought possible that processing might have a beneficial effect e.g. by gettering. This was the one of the motivations for the present study.

2.5 GETTERING

Gettering is the technique of trapping metallics in a region removed from the active device region. To do so requires the production of trapping sites, the diffusion of metallics to these sites and the formation of chemically stable structures at the trapping sites. Sites may be introduced intrinsically or extrinsically. Extrinsic sites are produced by mechanical damage, ion implantation or polysilicon deposition. Intrinsic sites are produced either by a region of SiO_2 precipitates (this is well discussed by Ping, Wang et al 1984 and Craven 1985), via the thermal introduction of Si

interstitials (Nauka et al 1985) or heavy B or P doping. Polignano et al 1988 have shown that As and Sb do not have a gettering effect and that B and P doping is more effective than SiO_2. They also suggest that the gettering process is not a strong function of the concentration of extended defects. The mechanisms of transition metal diffusion are dependent on the metal species, although many diffuse via interstitially. At the getter site, again the trapping mechanisms vary - Au is substitutional in P doped layers (Lescronier et al 1981) and Ni and Fe are interstitial and form $NiSi_2$ or $FeSi_2$ precipitates. Bailey et al 1985 have shown C gettering by dislocation and stacking faults. The effectiveness of gettering has been demonstrated by Borland 1985, who achieved a three order of magnitude increase of generation lifetime in CVD epilayers. A brief review of gettering in Si is given by Sadana, 1985.

The problem remaining for MBE is that these processes all involve long, high temperature anneals. Whilst getter sites may be induced in the substrate before epitaxy, the metals must be diffused to them and trapped. High growth, or post growth processing, temperatures are not compatible with the retention of abrupt dopant profiles or strained alloy structures grown by MBE. One option is rapid thermal processing (RTP). Sparks et al 1986 applied RTP and mechanical back damage, using temperatures of 300 to 1100 C and times of 100 to 300 seconds, all of which produced anomalous rapid gettering behaviour. Unfortunately, RTP is also known to produce deep levels: Tokuda et al 1989 suggest that RTP can introduce trap concentrations of 10^{13} cm^{-3} by thermal stressing: Eichhammer et al 1989 and Fang Lu et al 1992 have seen a large degradation in recombination lifetime with RTP and attribute this in part to the role of transition metals. This technique may yet prove to be useful, but there remains a problem specific to MBE (although it can exist, to a lesser extent, in other epitaxial techniques) that renders it presently unpromising as a general solution.

This MBE specific problem, involves the epilayer-substrate interface, which acts as a very effective metal trap site. This is most easily demonstrated using Cu,

which has a very high diffusion coefficient in Si, Fig. 2.8. Fig 2.9 shows the Secondary Ion Mass Spectrometry (SIMS) profile of a layer containing Cu before and after annealing. It is not known, conclusively, whether this gettering effect is due to the B interface spike rather any other property of the interface, i.e. residual SiO_2. However, it is clear that the combination of high dislocation levels (10^3 cm^{-3} in the sample of Fig. 2.9) and the interface act to getter. This contaminated region will have a deleterious effect on many devices, unless it is several minority carrier diffusion lengths from the active device region, i.e. if very thick epilayers are grown. In addition, getter sites in the substrate are screened by this interface. Salih et al 1986 used CVD to grow Si on a SiGe buffer, thereby generating a misfit dislocation network, which acted to getter extrinsically the Si, and achieved a tenfold increase in generation lifetime. (Rozgonyi 1987 conducted a similar experiment). It should be noted that the intriguing possibility of gettering through the use of growth interrupts alone is suggested in Chapter Five.

For this study it is important to note that it is probable that T_s plays a role in intrinsic gettering to the interface, that this interface is likely to be of very poor electrical quality and that the role of the interface may complicate a substrate gettering procedure.

2.6 SUMMARY

The role of unintentional impurities in MBE Si has been reviewed with relation to intentional doping in MBE. The complexity and multiplicity of the potential origins of electrically active point defects is such that few mechanisms for the degradation of material quality with T_s can be proposed with confidence. The concept of gettering has been introduced, and it is clear that, if gettering is found to be necessary in MBE

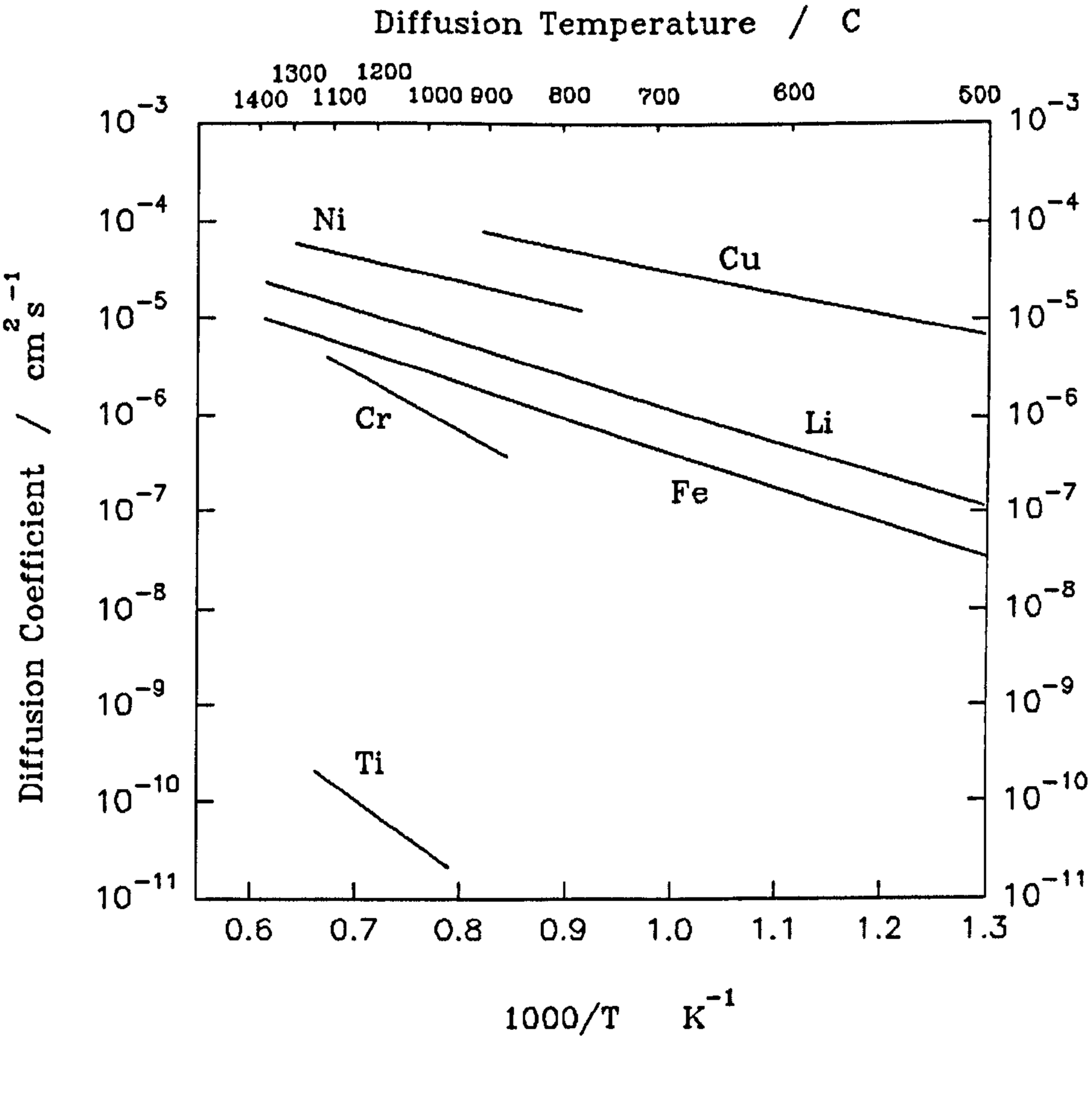


Fig. 2.8 Diffusion coefficients of 3d transition metals in Si at Si MBE growth temperatures (where known)
(After Weber, 1980)

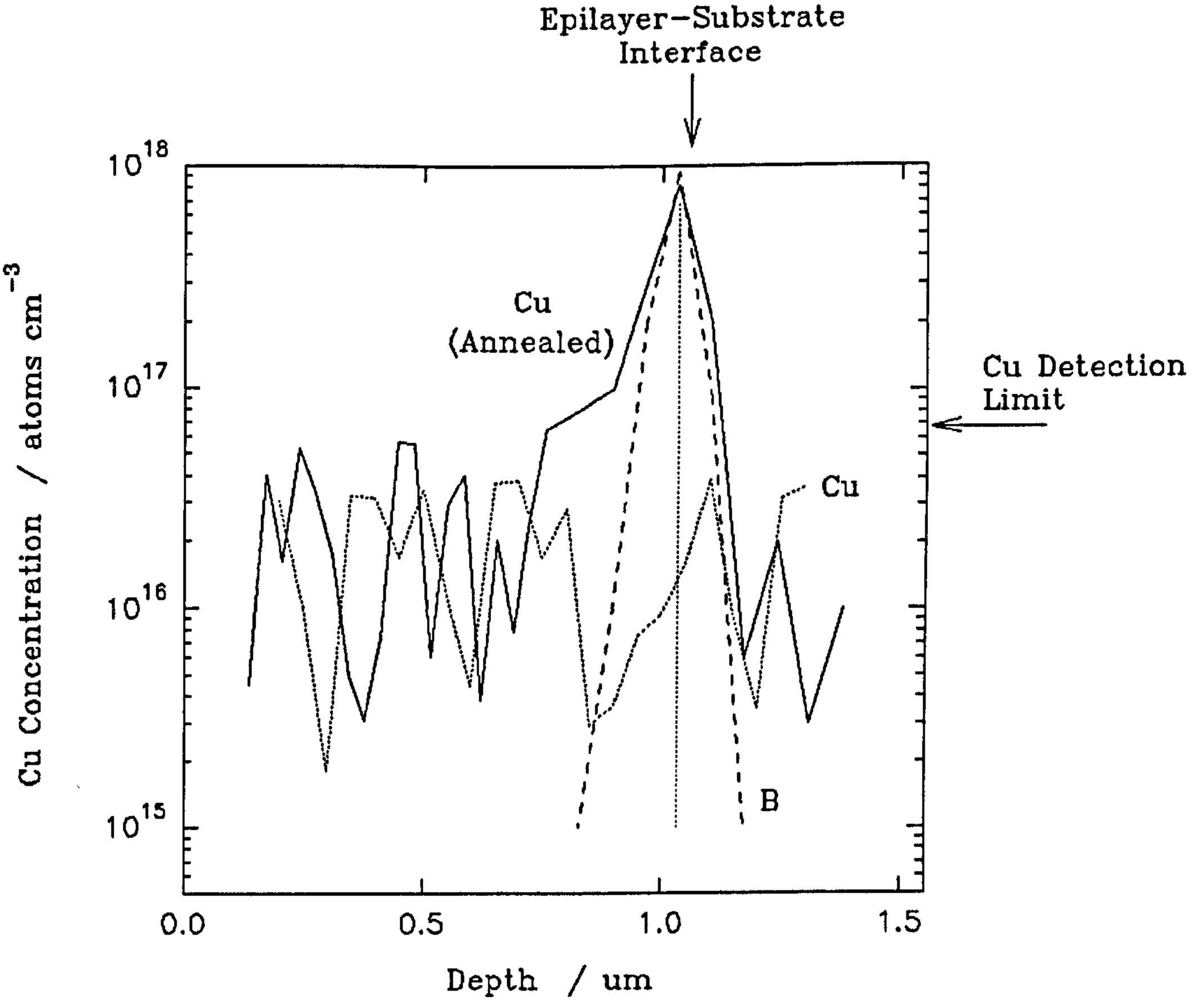


Fig 2.9 SIMS depth profile of a Si epilayer, before and after annealing at 1000 C for one minute. (After Houghton 1991)

material, which is probable (see Chapter Four), the lengthy, high temperature techniques often currently used for bulk material are unlikely to be directly applicable.

CHAPTER THREE

DEVELOPMENT OF HIGH INTEGRITY PROCESSING

3.1 INTRODUCTION

In order to produce a measurable structure it is necessary to define its geometry and to contact to selected regions. In many cases, it is also necessary to passivate the semiconductor surface and to produce a means of external modulation of the charge carriers in the structures. For many simple measurements (e.g. the Hall effect), it is possible to achieve this with relatively simple equipment - "in-house" processing. For others, the techniques of VLSI Si processing are imperative. Further, it is important to establish the effect of such processing on MBE Si and to develop process steps as necessary, that do not have a deleterious effect on the device. On the initiation of the author, VLSI processing was accessed at the Edinburgh Microfabrication Facility (EMF), and low temperature processing steps developed by the staff there, in conjunction with the author. Use was also made of the Southampton University Centre of Electronic Materials (SUCEM), for whom the author developed a self-aligned gated Hall bar process.

These sections will give experimental details for both the "in-house" and VLSI processing that have been used in this study. Minimal details are given of the conventional EMF techniques employed, but comprehensive discussions are found in many books e.g. VLSI Technology by Sze. The theory necessary to explain the experimental approach and results is given in Section 3.5. It is stressed that this work represents a 'means-to-an-end' for the growth and characterisation of Si MBE devices.

While the author attempted to avoid developing unestablished procedures at all times, this did become necessary on a number of occasions, as detailed in the following.

3.2 THERMAL BUDGETS

It has been fortunate that the need to reduce temperature/time cycling for MBE Si has coincided with the same need in conventional processing, primarily to reduce diffusion effects. These effects assume significance as packing densities increase and minimum feature sizes decrease in order to improve device speed and yields. The term 'thermal budget' is commonly used to refer to the temperature/time cycling of a given process step. This has been defined as $\sqrt{Dt}$, where is D is the diffusion coefficient at a given temperature for a given species, t is time; ($\sqrt{Dt}$ is, of course, the mean distance travelled through a medium in time t). This definition, however, implies different thermal budgets for each species (e.g. dopant) present in a given process step. (In the universal sense that thermal budget is often used, a definition of thermal budget

$$= \int_0^{t_f} Tt.dt$$ might be more appropriate.)

It is necessary to ask what the maximum permissible thermal budget is for MBE grown material. In order to minimise process development and assessment, this question must be answered for the most temperature sensitive structures (i.e. δ layers and strained SiGe), from which the general prescription must be based. In the absence of other information, the aim at the start of this study was to keep process temperatures below T_s i.e. <550 C. The exception to this prescription was an implantation anneal discussed in Section 3.6.2. Since post growth annealing is a pseudo-equilibrium process and MBE growth is not, this might appear naive and, in the course of this study, it has become clear that grown structures are stable to significantly higher anneal

temperatures. Diffusion, however, is a multi-parameter phenomenon, highly sensitive to imperfections in the host lattice and MBE Si is subject to higher point and extended defect concentrations than substrate material. The role of defects in diffusion mechanisms is well reviewed by Fahey et al 1989.

During the course of this work, X-ray diffraction (for a review see Tanner & Bowen 1993) double crystal rocking curve and SIMS studies have yielded information regarding maximum permissible anneals for δ layers (Powell et al, 1991a) and metastable SiGe (Powell et al 1991b). These indicate that one hour anneals at 700 C broaden a delta layer FWHM from 0.3 nm to 4.5 nm, which is acceptable. However, at 750 C, the FWHM is 15 nm, which is not (Fig 3.1); also, that strained SiGe, up to three times the equilbrium t_c, can withstand 850 C for one hour before unacceptable relaxation occurs. The effects of such anneals on midgap states was unreported. The author's work in this area is discussed in Chapter Four.

3.3 MASK DESIGN

In order to define structures using a photoresist, it is necessary to produce optical masks. (The author has designed all of the masks sets for in-house Hall bar manufacture, for gated Hall bar manufacture at the SUCEM and for many structures at the EMF.) Mask design is governed by the nature of the process steps used and their order. Thus, the exact process for a given structure must be determined before the masks are laid out, so the mask designer is, of necessity, also the process designer. This is less true of standard MOS or bipolar processing where design rules are available and often incorporated into design software or where process modifications can be dictated to the mask designer. However, for processing the structures of this study, these were not options.

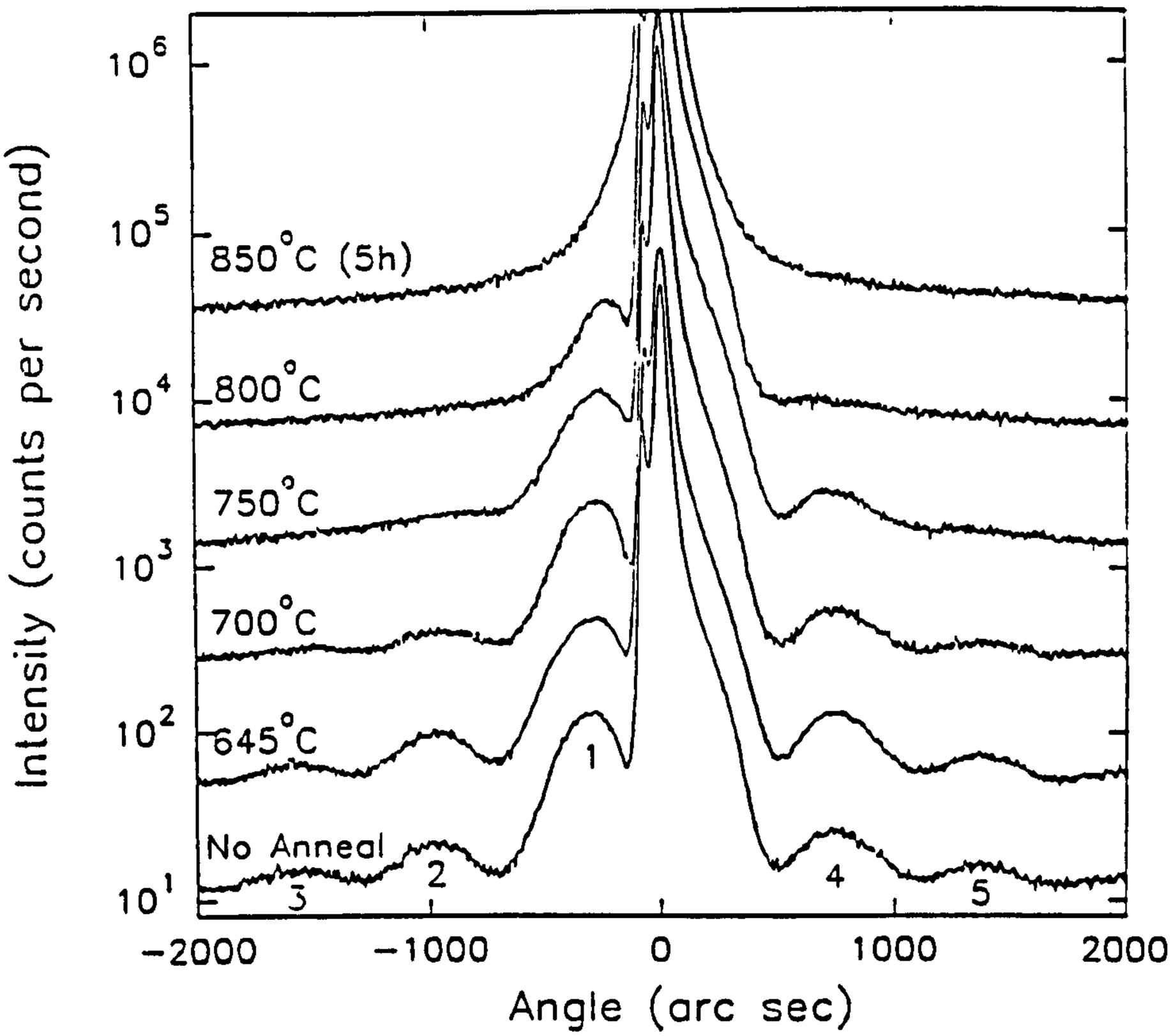


Fig. 3.1 X-ray diffraction results showing the effect of one hour anneals on a 3.6×10^{14} cm^{-2} B delta layer, with significant broadening at temperatures above 700 C. (After Powell et al 1991)

Masks for in-house work were designed using a general software package, ECAD: these designs were laser printed, photographed and scaled down by factor 10. At Warwick, the final image, which was double reversed, was developed on glass plates coated with photographic emulsion. These were then used in a UV contact printer with 50 μm minimum feature size. For processes at SUCEM, the same design process was used but negative 35mm film was supplied to SUCEM, where the image was transferred to 4" quartz plates for use in a IR contact printer that is normally used for double sided alignment. Mask sets for the EMF were designed initially using software obtained from the Rutherford Appleton Laboratory (RAL) PRIME network - Graphic Aided Engineering Layout of Integrated Circuits (GAELIC). For the subsequent two mask sets, California Intermediate Form (CIF) was used. Files containing these designs were sent to the RAL, where global alignment marks were added. The masks were then brokered out for making by e-beam definition of Cr on 4" quartz plates. The minimum feature size of the EMF processes was 5μm, which corresponds to ± 0.25 μm critical dimension tolerance on the mask.

The mask sets for in-house processing are the most straightforward, consisting of 2 light field masks, the first for mesa definition, the second for Al contacts. A very similar process was developed for SUCEM with the addition of a Schottky gate: this required an additional dark field plate for a metallisation lif-off, self-aligned gate process shown in Fig.3.2. It should be noted that both the EMF and SUCEM use negative photoresist only: at Warwick, the author used both negative and positive where appropriate.

At the EMF, a divide by 10 optical step and repeat Optimetrix projection system is used for patterning and mask alignment. Although the main use of such projection printing is its tolerance (0.1 μm), a further advantage here is that a single design may be positioned anywhere on the wafer such that the area to be processed may be defined at the time of processing rather than when mask designing. Chips from the EMF were

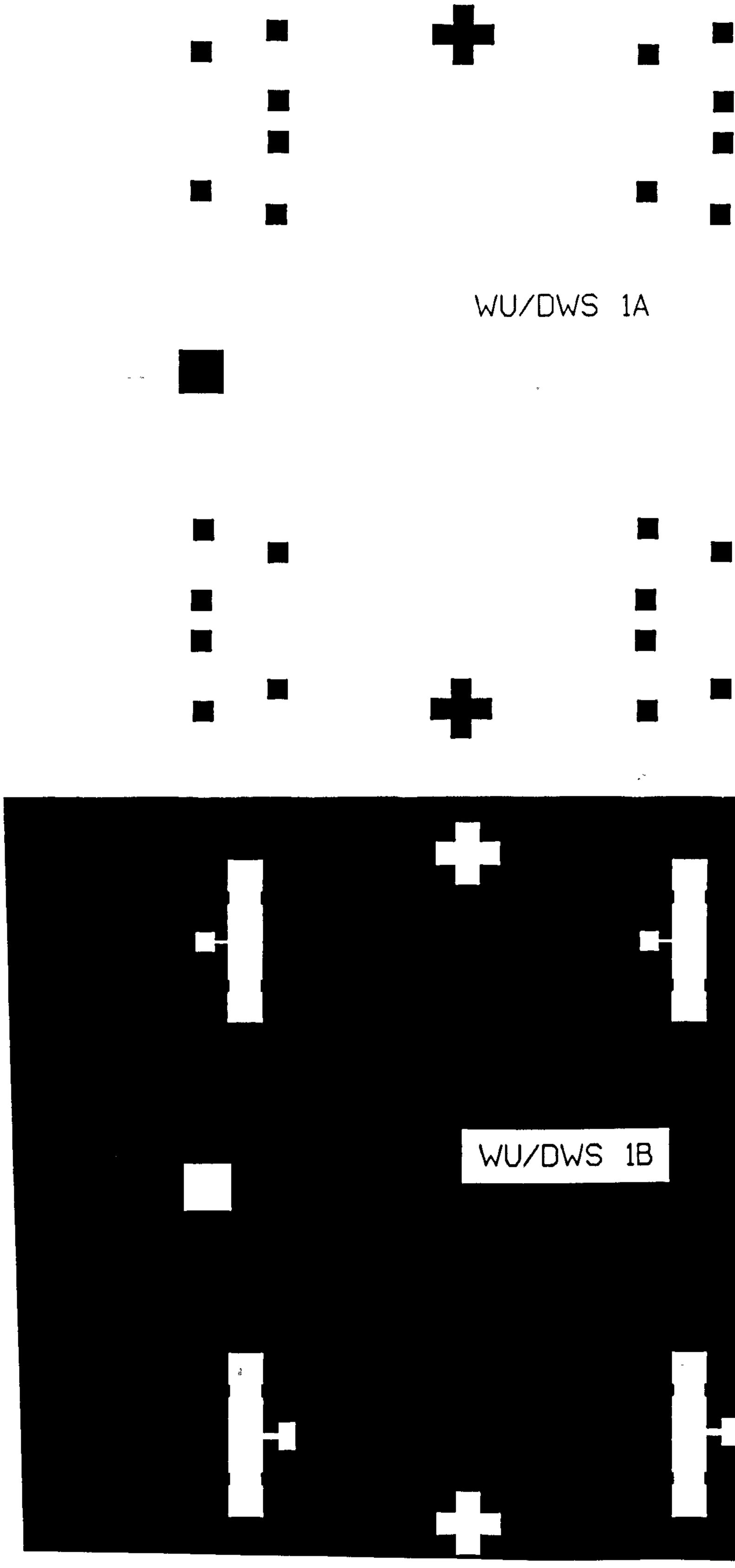
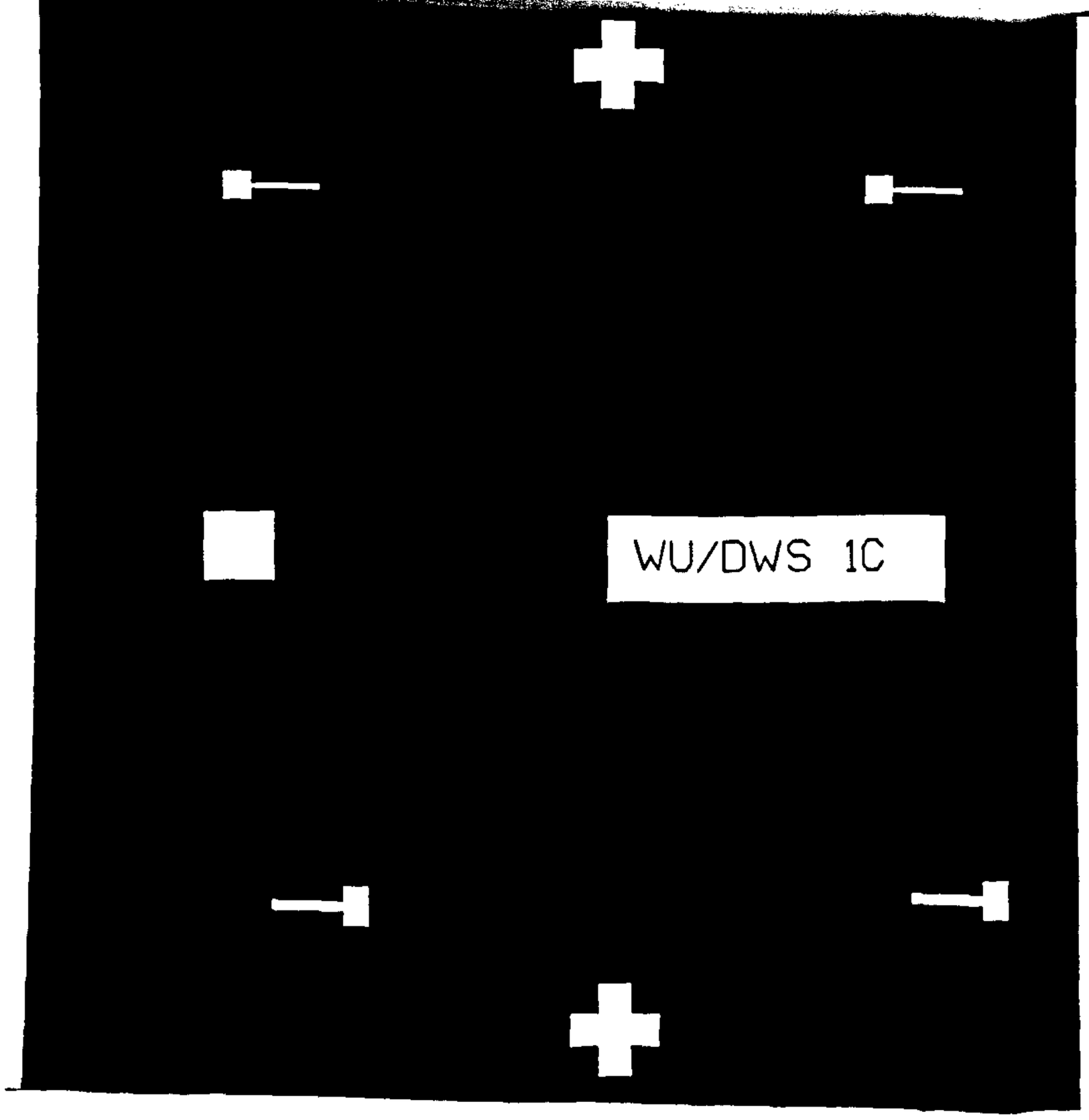
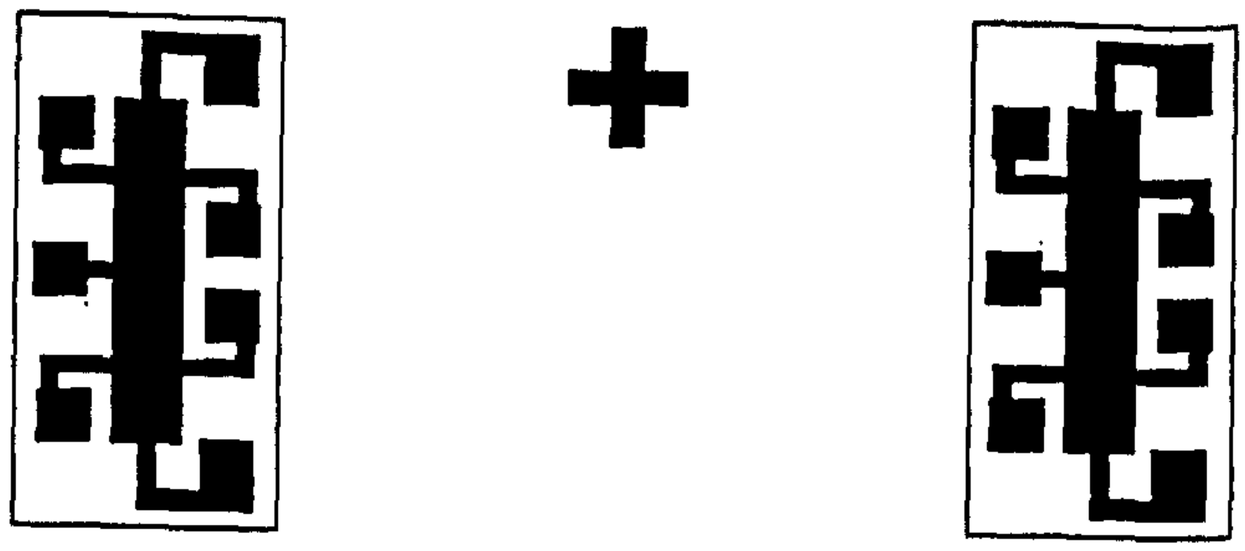


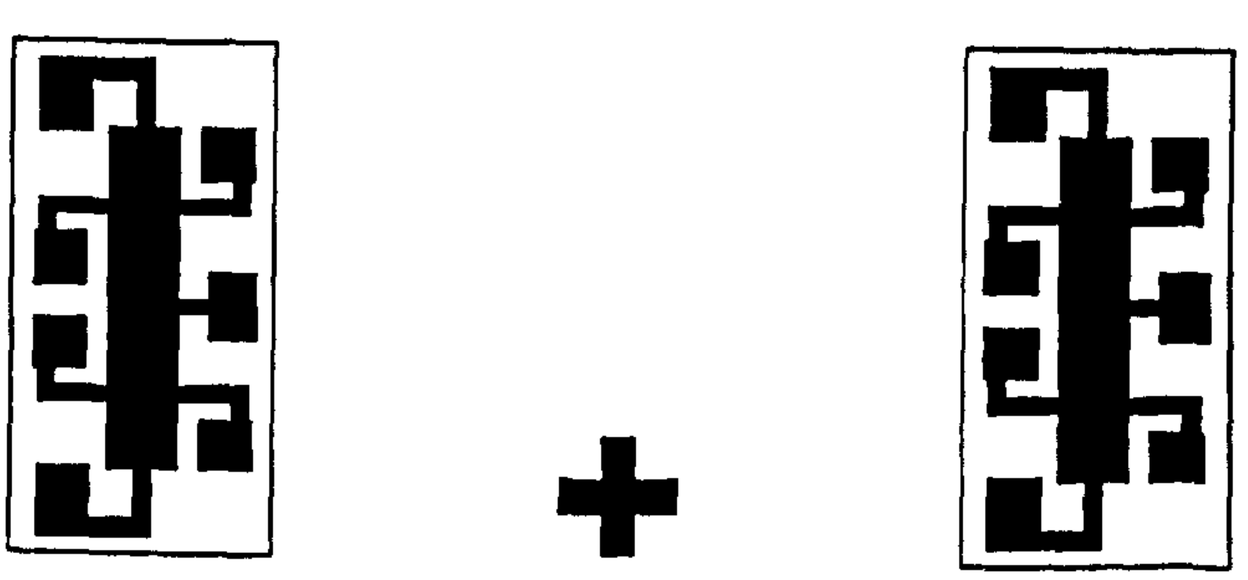
Fig. 3.2 Hall bar masks for self alligned Schottky gate process. (a) Al contact definition, (b) and (c) optional Shottky gate definition masks, (d) mesa etch mask.



WU/DWS 1C



WU/DWS 1D



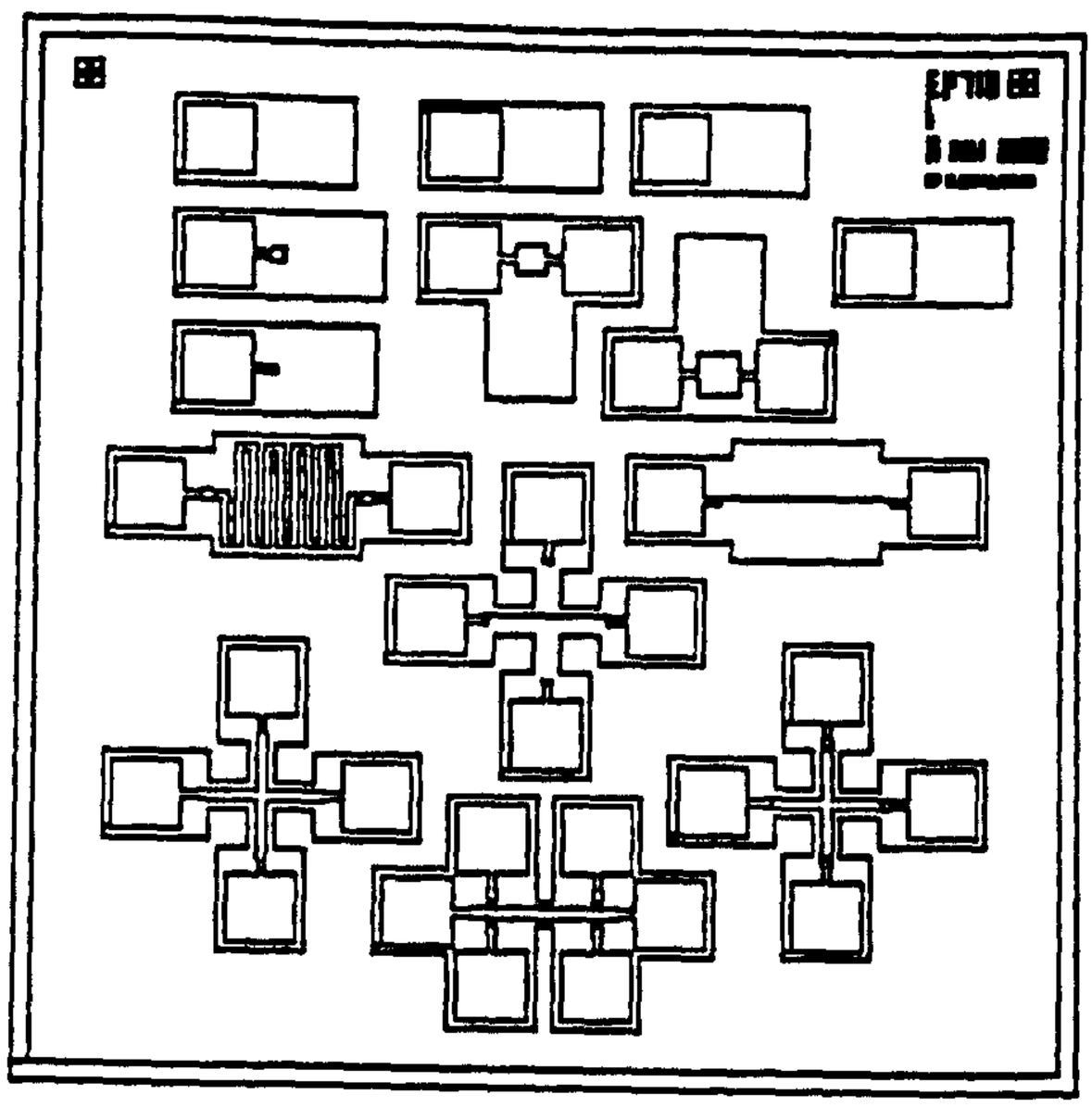
laid down in either chequerboard formations or in alternate rows: this leaves unprocessed material adjacent to each active chip for analysis via SIMS, e-CV etc. This was particularly useful for V80 material when Sb doping levels could vary by one or two orders of magnitude across a wafer. Doping and alloy uniformity for the V90S system has been found by X-ray diffraction to be ± 1% across a 4" wafer. A 3mm x 3mm frame (i.e. 'chip') was chosen as the largest size that, when packaged, would allow access to all of the available cryostats. This requirement became less stringent later in the study, so the in-house and SUCEM frames were chosen to be 6mm x 3mm to enhance yield, thereby increasing the number of Hall bars available per process. The chips were packaged and bonded using a thermo-compression Kulike and Sofa 4124 gold ball bonder.

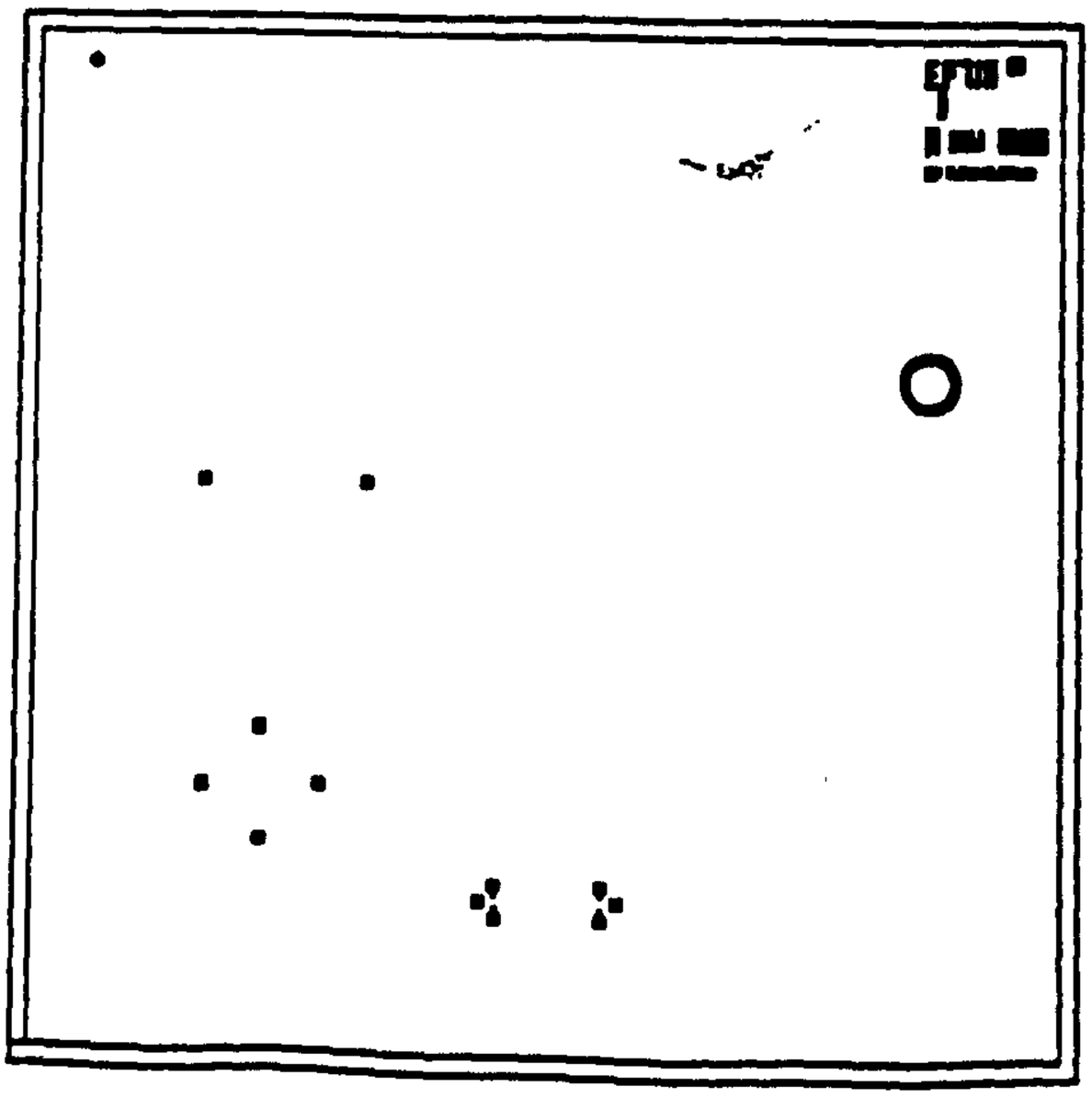
The general design strategy was to minimise the number of process steps for each structure. It was not practicable, nor desirable, to produce chips such that each structure type was functional for a given process owing to the diversity of epitaxial structures for which each mask set was designed. However, where possible, analytical structures such as Hall bars were made coherent with device structures.

There are 30 different structures on the 3 mask sets so it is not sensible to list each structure and the relevant process steps. A key to the final structure is given on the first two mask sets shown in Figs 3.3, 3.4. In addition, each mask is labelled with the use to which it was put most commonly, notwithstanding that what is nominally an oxide window mask might act as an implant mask for some structures. The final mask set is shown overlaid in Fig. 3.5 as an overview.

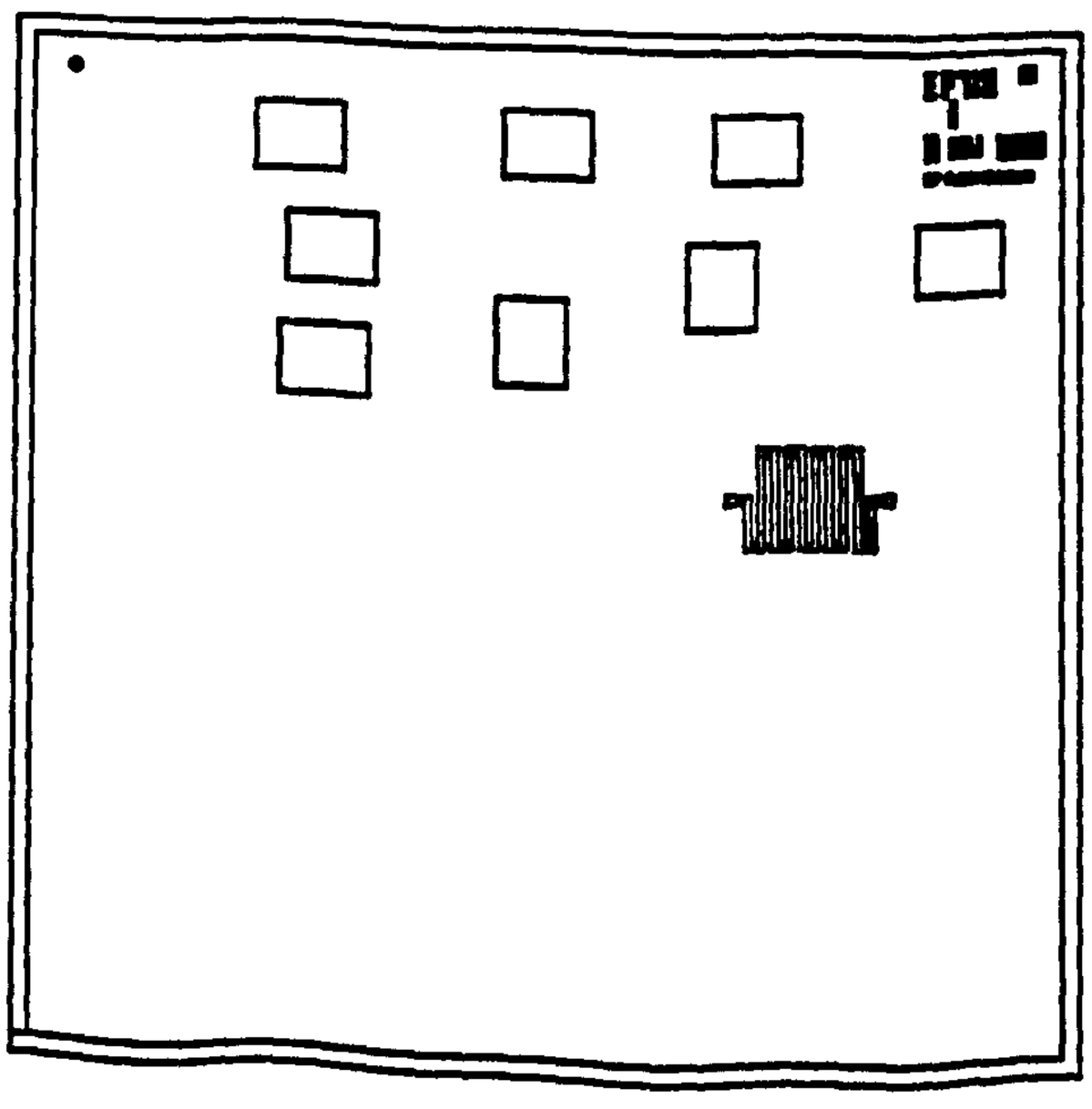
A schematic is shown, in Fig 3.6, of the process schedule of a δ FET in order to illustrate many of the processes used to define MBE grown structures and devices. Details of electrical results from the delta FET, selectively contacted n-i-p-i Si superlattice and gated Hall bars are found in Biswas, 1992. The author was solely

Mesa Etch





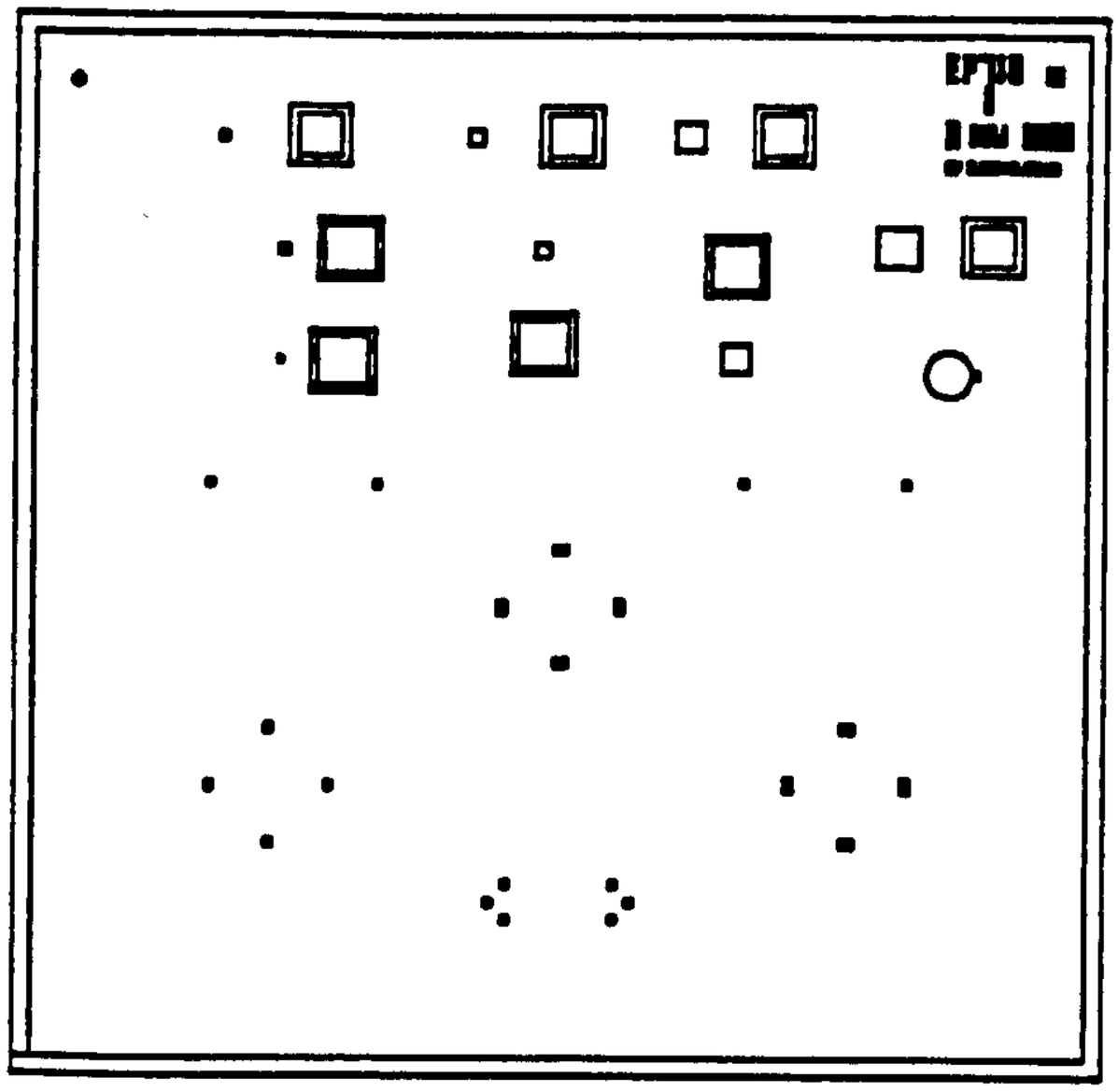
Implant 1.



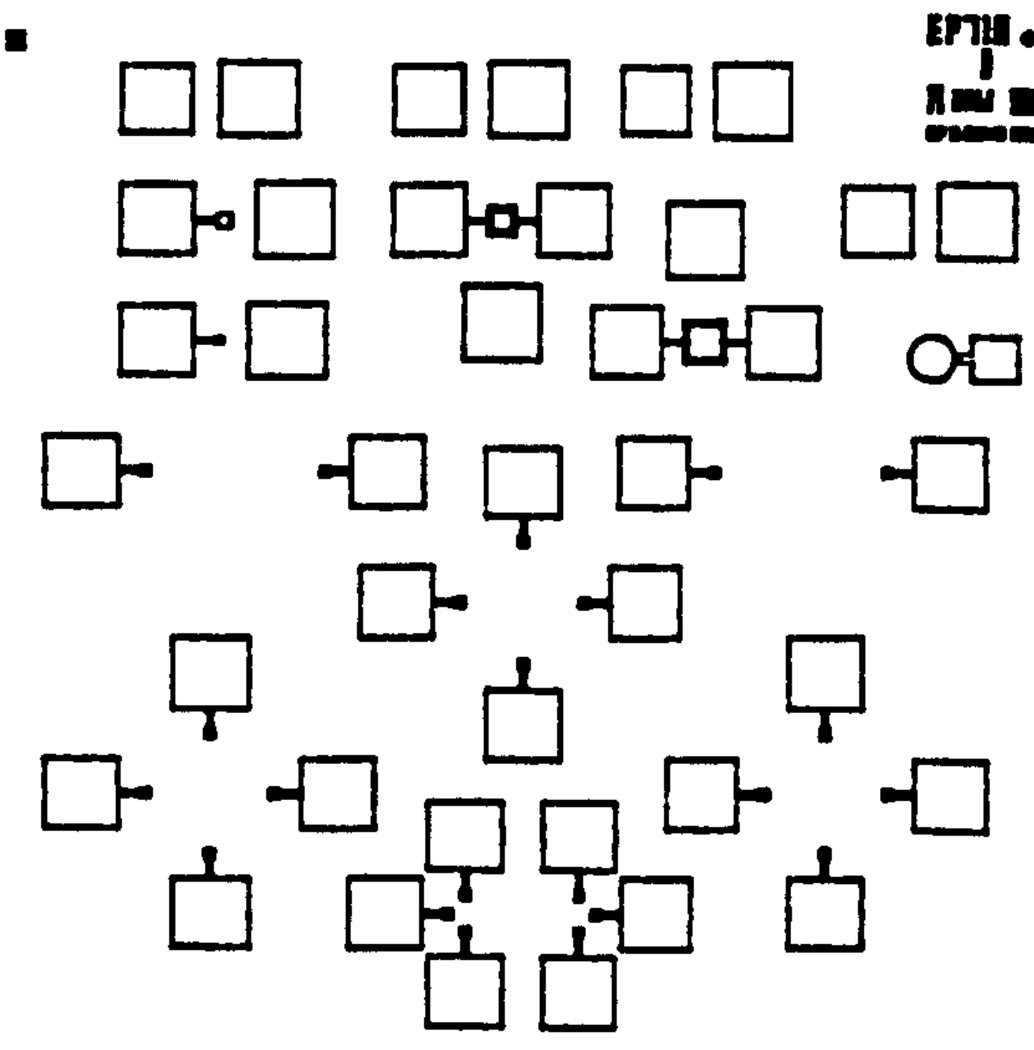
Implant 2.

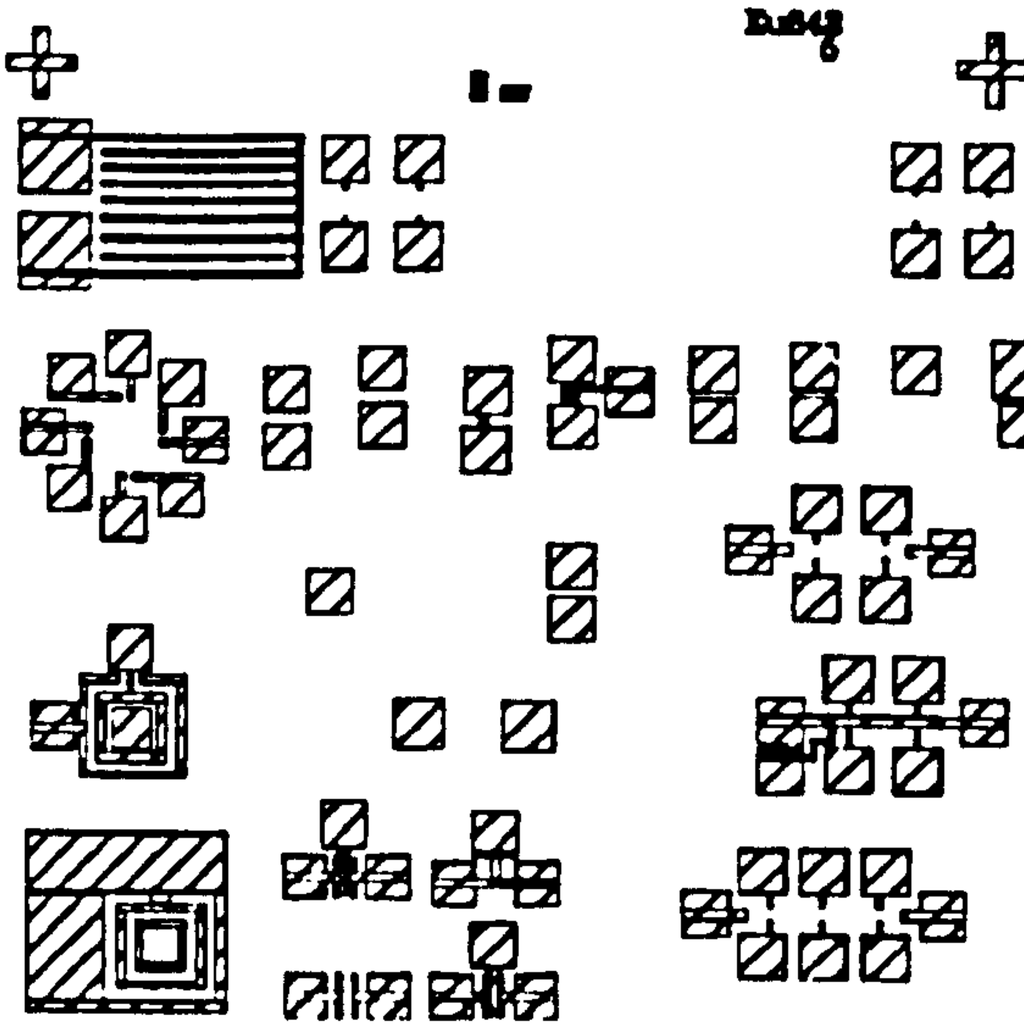
Fig. 3.3 Mask set E_μ 716.

Oxide Etch

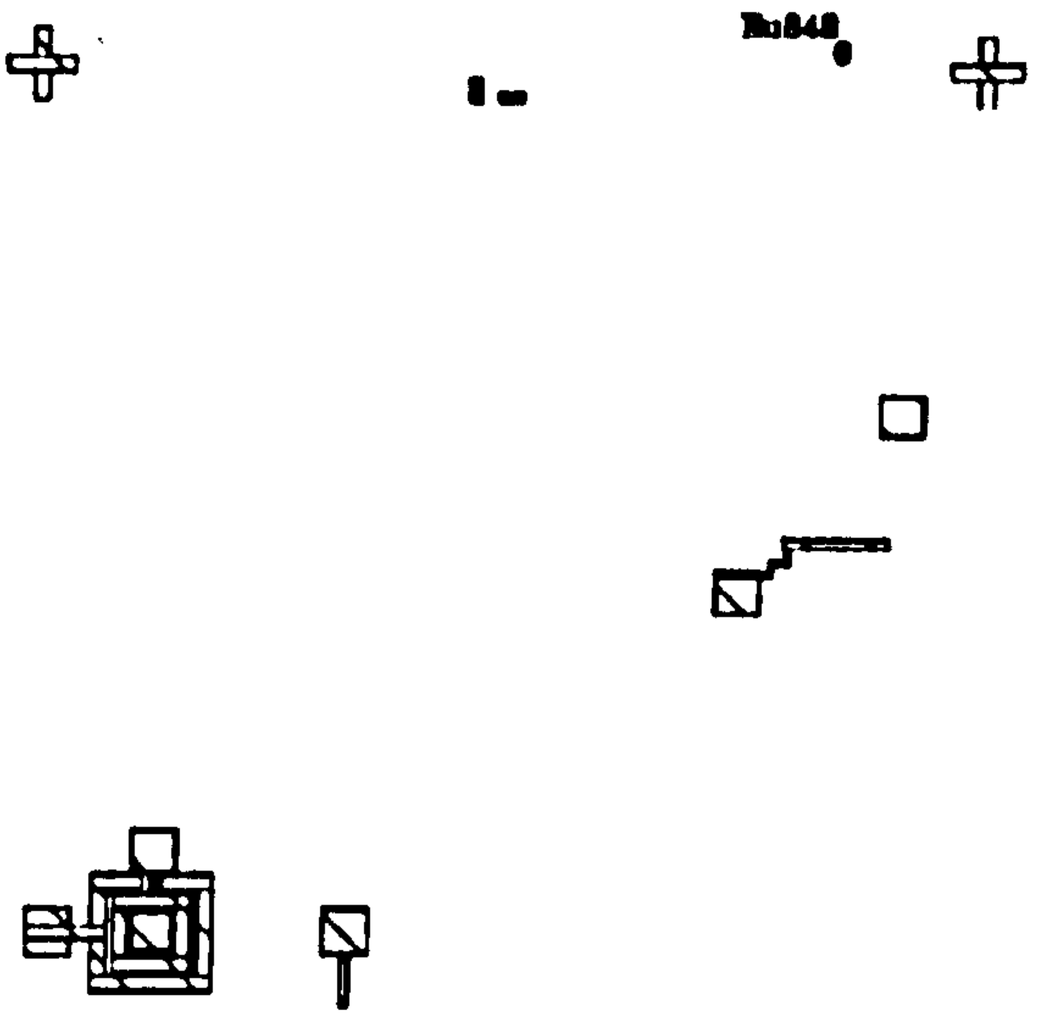


Metallisation

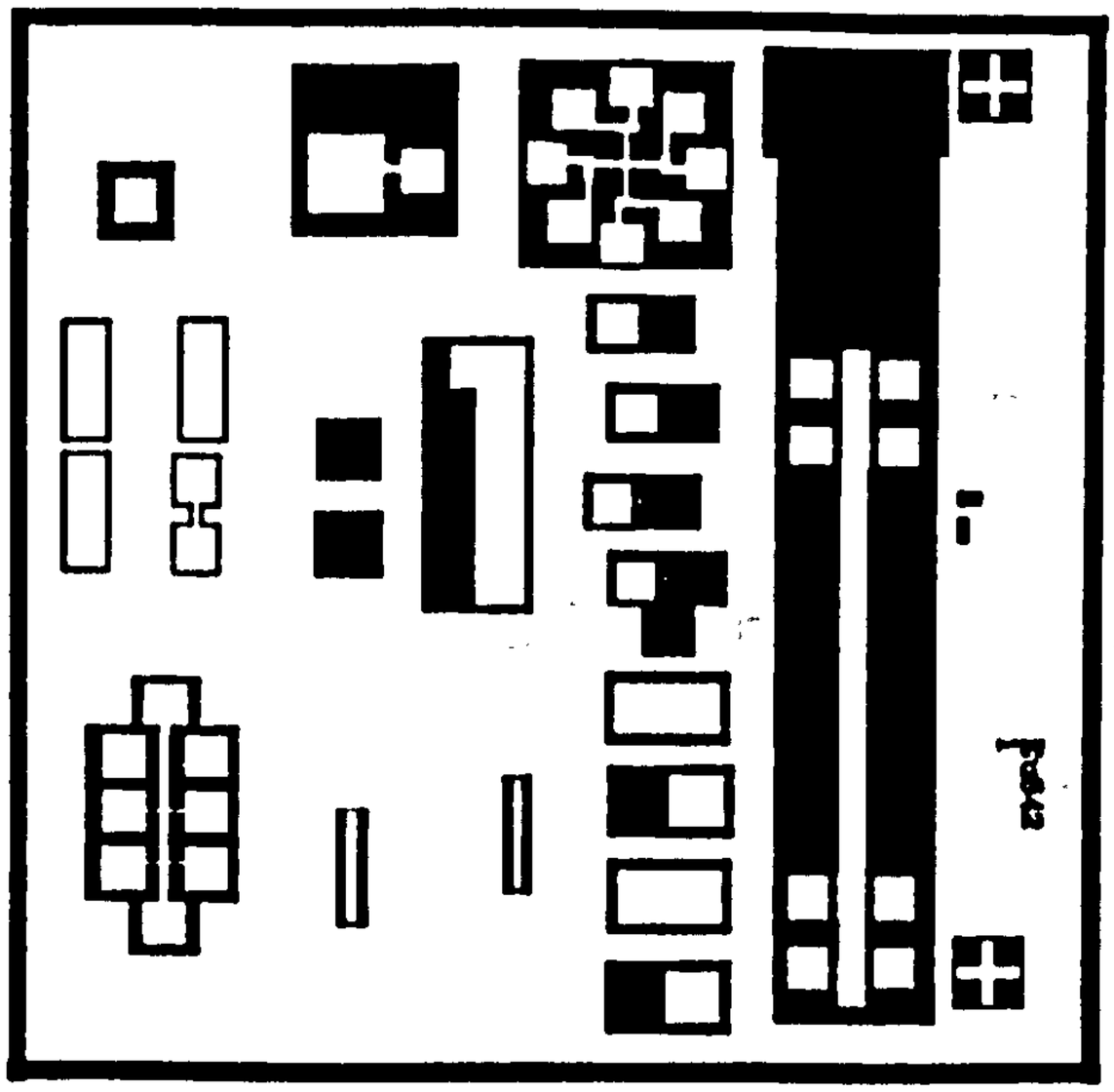




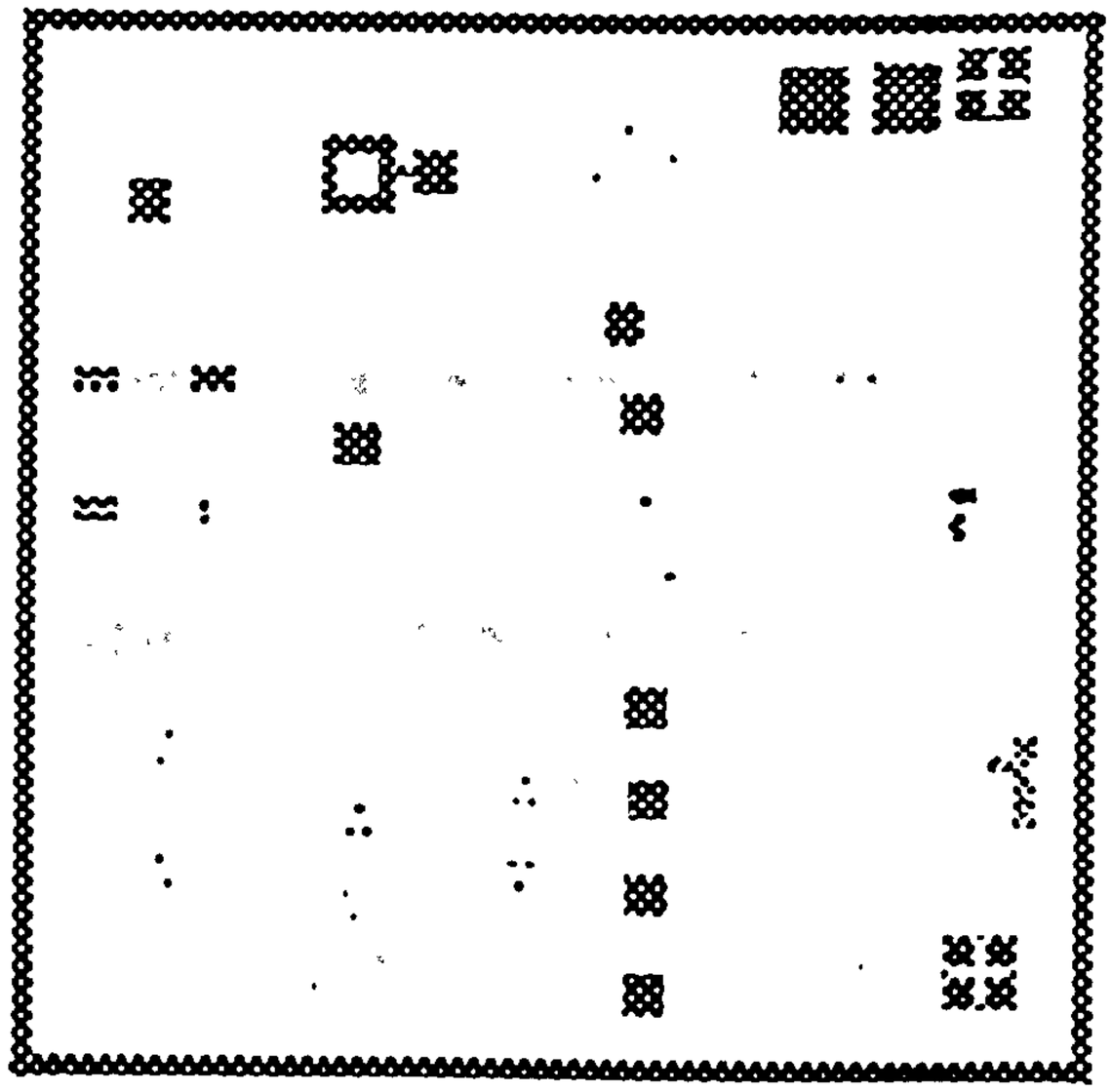
Metallisation 1.



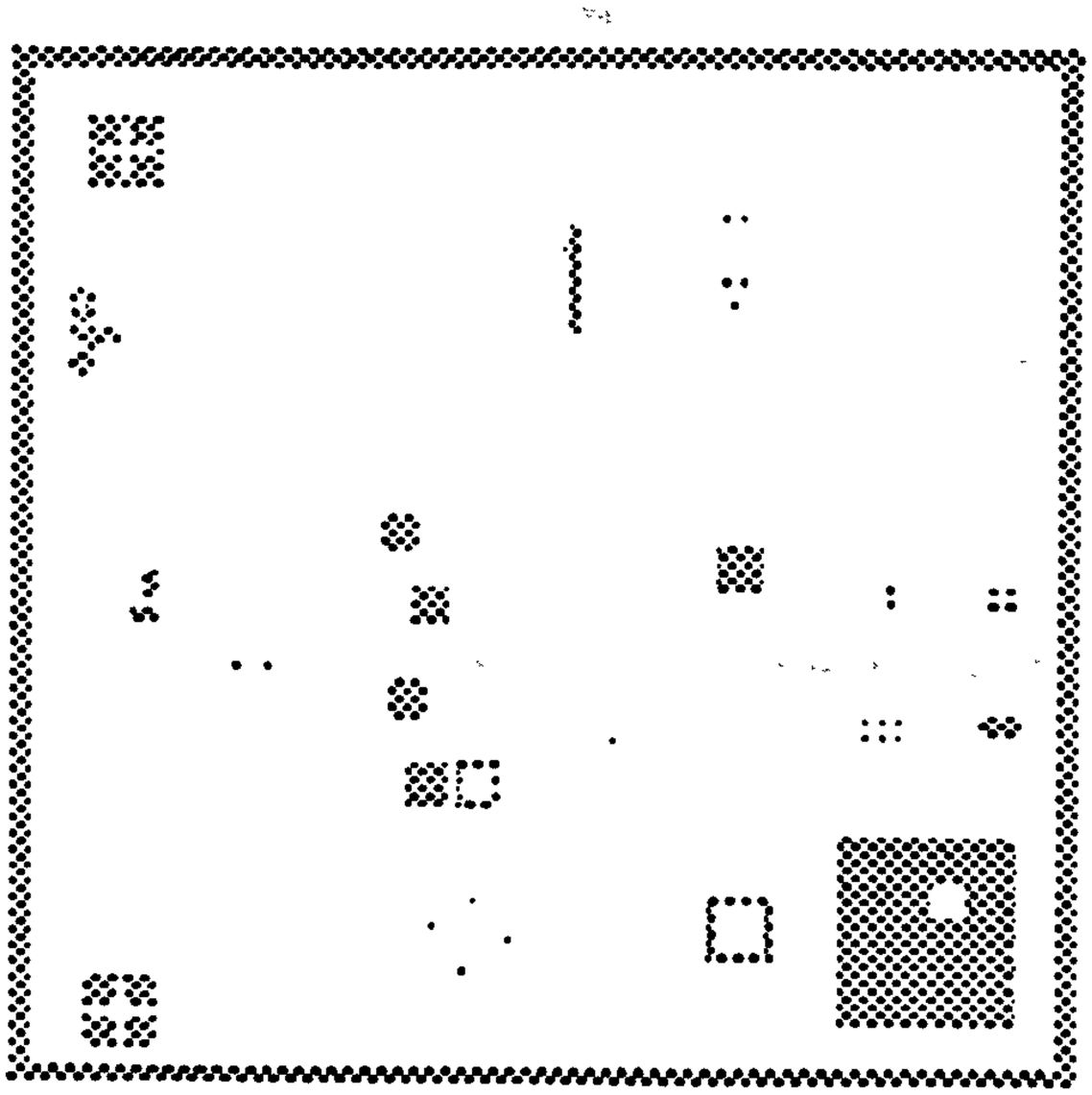
Metallisation 2.



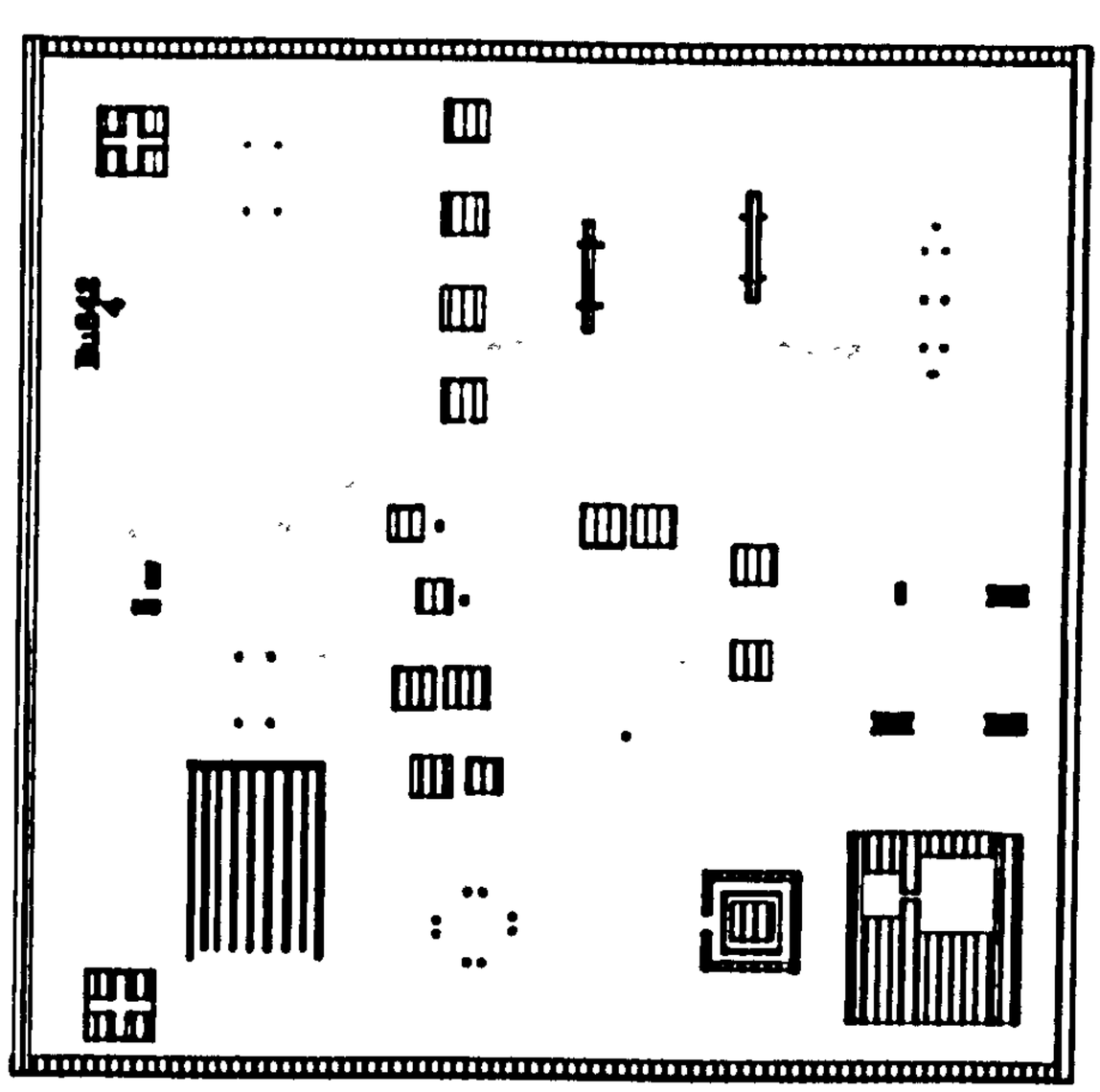
Mesa Etch



Implant 1.



Implant 2.



Oxide Etch

Fig. 3.4 Mask set E_μ 842.

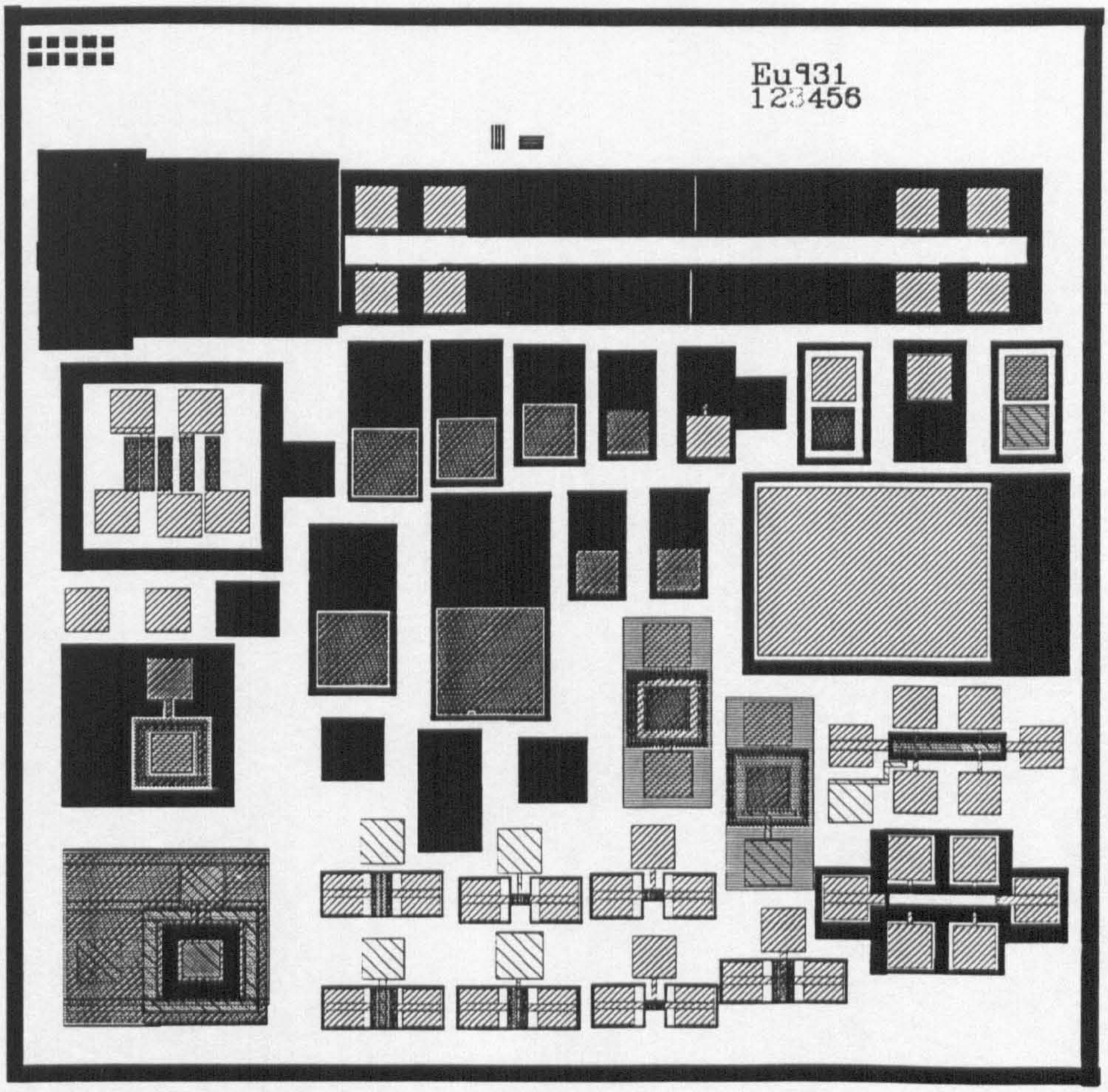
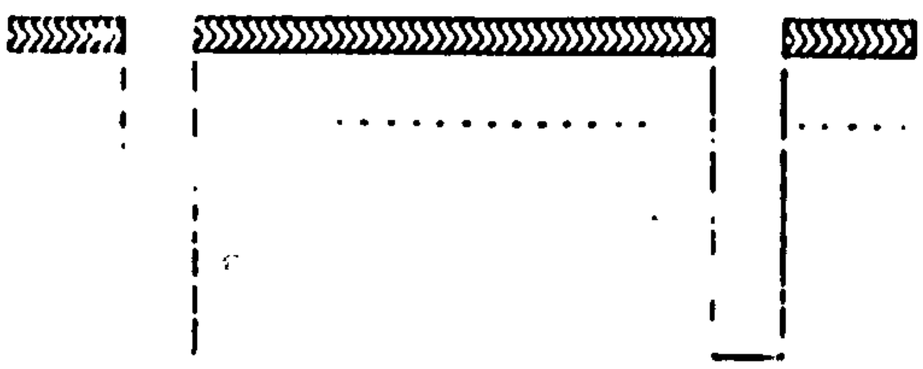


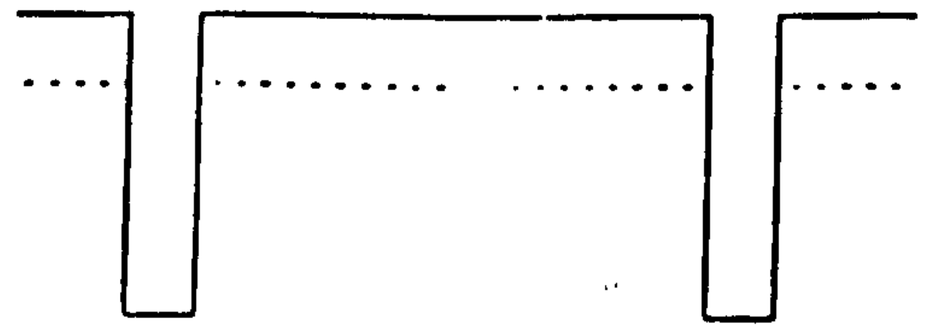
Fig. 3.5 Mask set E_{μ} 931.

(a) Pyrolytic oxide deposition (@.430°C).

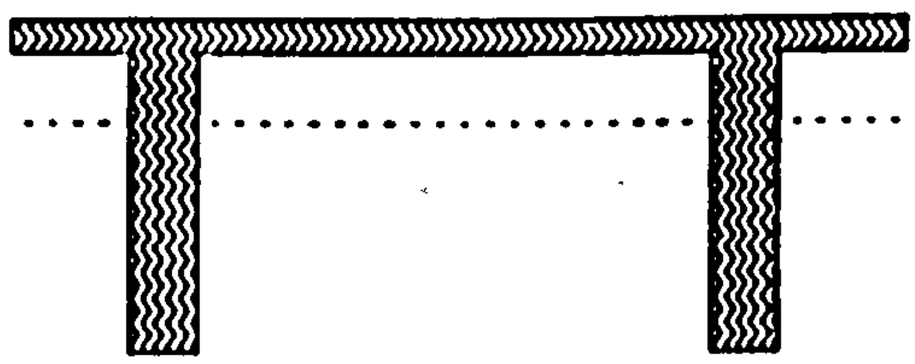
(b) Oxide reactive ion etch (RIE).



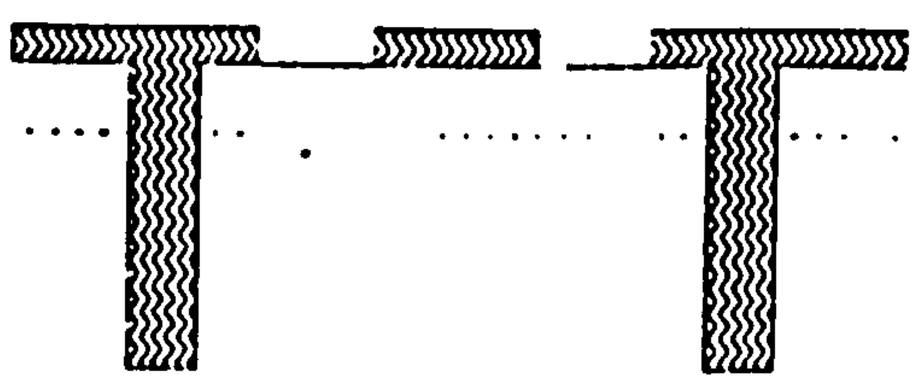
(c) Trench RIE etch.



(d) Oxide strip.

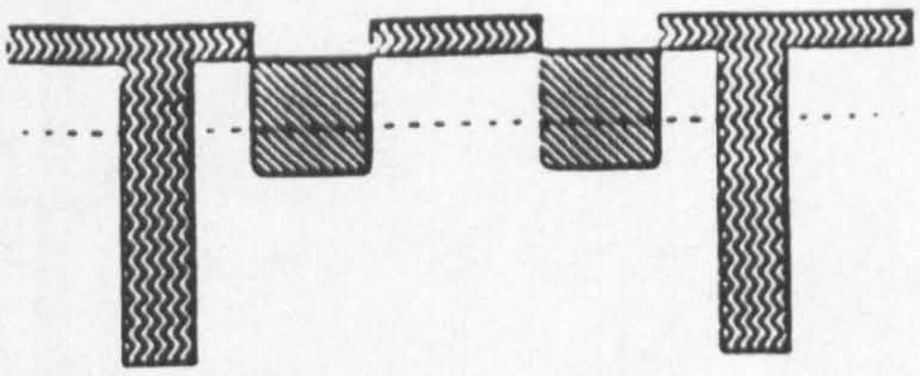


(e) Pyrolytic deposition.

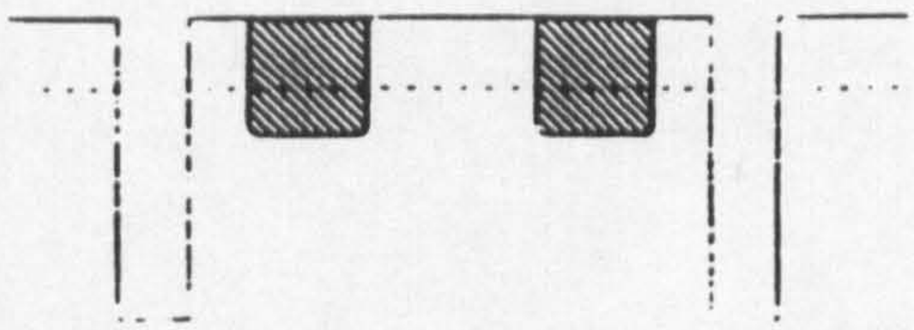


(f) Oxide RIE etch.

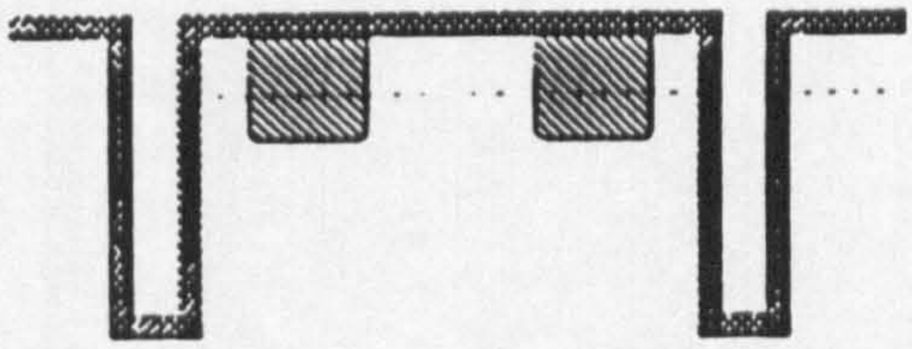
Fig. 3.6 The process sequence for the low thermal budget fabrication of an MBE grown B δ FET - continued overleaf. (Diagrams after Biswas, 1991, process and mask design by the present author.)



(g) Contact implantation
(BF_2 $1x10^{15}$ atoms cm^{-2} @ 40keV).

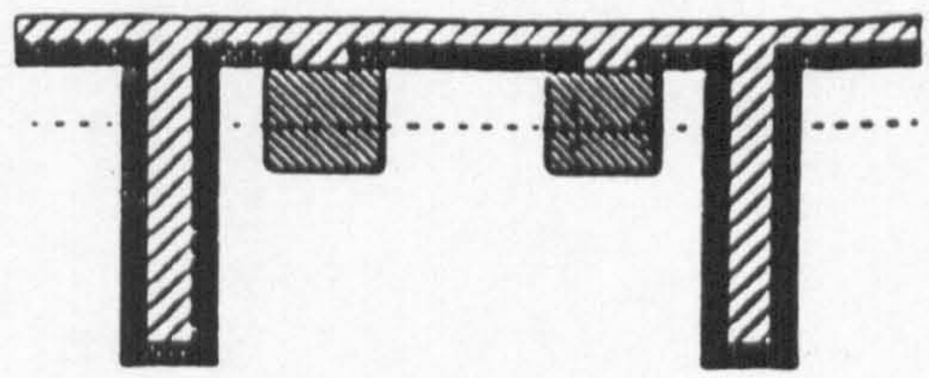


(h) Oxide strip.

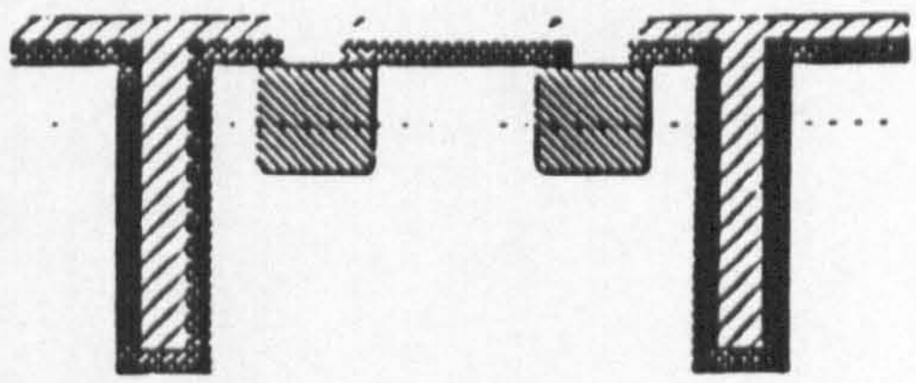


(i) Low temperature plasma oxidation.

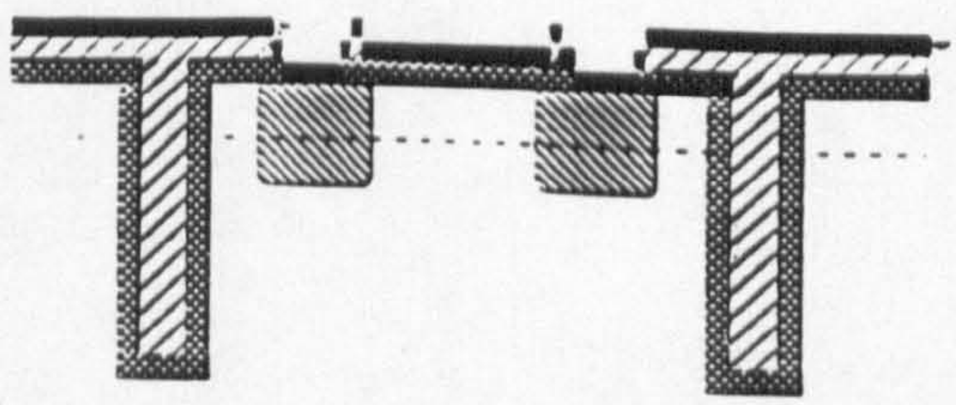
(j) Oxide etch.



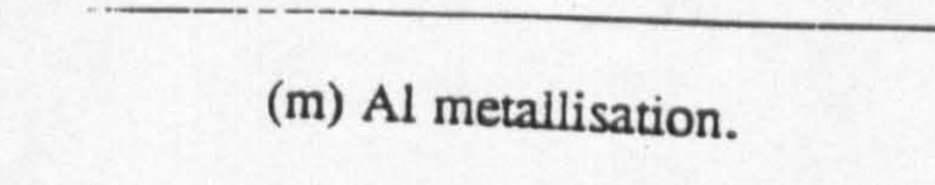
(k) Rapid thermal anneal (1000° 15s),
Polyimide coat and cure.



(l) RIE for contact windows and gate exposure



(m) Al metallisation.



(n) Al anneal (435°C 20 minutes).

responsible for the design of all of these structures and process steps with the exception of the delta FETs where N L Mattey gave assistance.

3.4 MESA ETCHING

3.4.1 Wet Etching

To define structure geometry or to isolate the planar doped structures inherent in MBE, mesa structures are created by etching. "In-house", this is achieved by masking with 'Apezion W' black wax, etching in HNO_3:HF:Acetic (5:3:3) - the CP4A etch - which has an etch rate of 0.1um /sec - and wax removal by xylene. The etch reaction is complex, but the nitric acid forms SiO_2 which is then etched by the hydroflouric acid. The acetic acid moderates the reaction. The limitations of this procedure are very poor control of etch depth, isotropy and a minimum mesa dimension of ~2 mm. This is of little significance for van der Pauw crosses used for Hall measurements, owing to the large size of the cross (3mm arm width,1cm arm lengths) and the use of four probe measurements that cancel asymmetry effects. "In-house" photolithography is used to produce Hall bars to ASTM standards. It is necessary to use a double barrier to the mesa etch, consisting of Al and resist, since conventional resists are not barriers to the Si etch. Both Al and resist are attacked by the CP4A etch; therefore, this technique does not rely upon a high ratio of Si to barrier etch rate (the ratio is approximately unity) but on a barrier thickness much greater than the mesa required in the Si. For most practical structures this is readily achieved, since the conducting layer(s) requiring isolation is within 0.5 μm of the surface.

3.4.2 RIE Etching

At the EMF, reactive ion etching (RIE) is used for its anisotropic etch, consistent and controllable etch rate and greater cleanliness (see for example Downey et al 1981). The reactor strikes a plasma by means of 2 parallel plate powered electrodes, on one of which the wafers are mounted. The chamber is of Al construction, is vacuum pumped and is a batch system. For Si etching, the reactant molecular gas was BCl_2. Etching occurs by dissociation, adsorption, product formation and, finally, desorption, releasing $SiCl_4$: ion bombardment also induces sputtering.

The substrate only reaches temperatures of 150°C and no contamination has been reported as a result of RIE, possibly because it has been confined deep in the substrate. One problem remaining is a probable high density of surface states on the mesa sidewalls and crater bottom. Raman scattering studies by Tang et al, 1992, on SiGe quantum wires formed in material grown in the V90S, demonstrate considerably less sidewall damage than achieved in III-V structures.

In order to resolve some problems of contacting selectively to buried layers, a mass spectrometer was attached to the RIE system to enable end-point detection during the etch process. This provides a significant advantage as compared to wet etching and has been used by the author in the design of a process to contact separately to n and p-type δ layers separated by only 10 nm.

3.5 THEORY

Providing ohmic contacts to Si is a routine procedure, but an inexact science. The demands of some MBE grown structures require innovation and it is useful to review briefly the reasons why a metal/semiconductor contact may be ohmic or rectifying.

3.5 1 Metal / Semiconductor Contacts

Semiconductor surfaces are much studied, but the number of empirical recipes for contact production greatly outnumber those with well understood mechanisms. Any interface between two solids will give rise to a potential barrier, so a perfect ohmic contact, i.e. with perfect linear IV dependence, is not achievable. However, if the voltage drop across the contact resistance, R_c, $(=\frac{\partial J}{\partial V})$ is small compared to that across the bulk during normal operation, it may be considered ohmic.

The application of simply Schottky theory to metal semiconductor junctions yields a barrier height ϕ, with

$$\phi = \chi_m - \chi_s \qquad 3.1$$

where χ_m is the electron affinity of the metal (equivalent to the metal work function) and χ_s that of the semiconductor. However, as Crowell et al showed in 1965, measured barrier heights differ greatly from such predictions, (Fig. 3.7), even when image force lowering is considered. The primary reason for the failure of Schottky theory is the role of disorder at the interface. Sebestyen, 1982 suggests that current transport in this disordered region is due to variable range hopping between localised mid-gap states. Pre-metallisation surface treatment may, therefore, determine R_c: mechanical damage e.g. sandblasting, introduces a high density of recombination centres and has been used to lower R_c. There is another means of enhancing current transport which is commonly used, comprising high doping in the contact region. The effect of doping on R_c was examined by Yu, 1970. If we define the parameter

$$E_0 = \frac{q\mathrm{h}}{2}\sqrt{\frac{N_d}{m^* \varepsilon_r}} \qquad 3.2$$

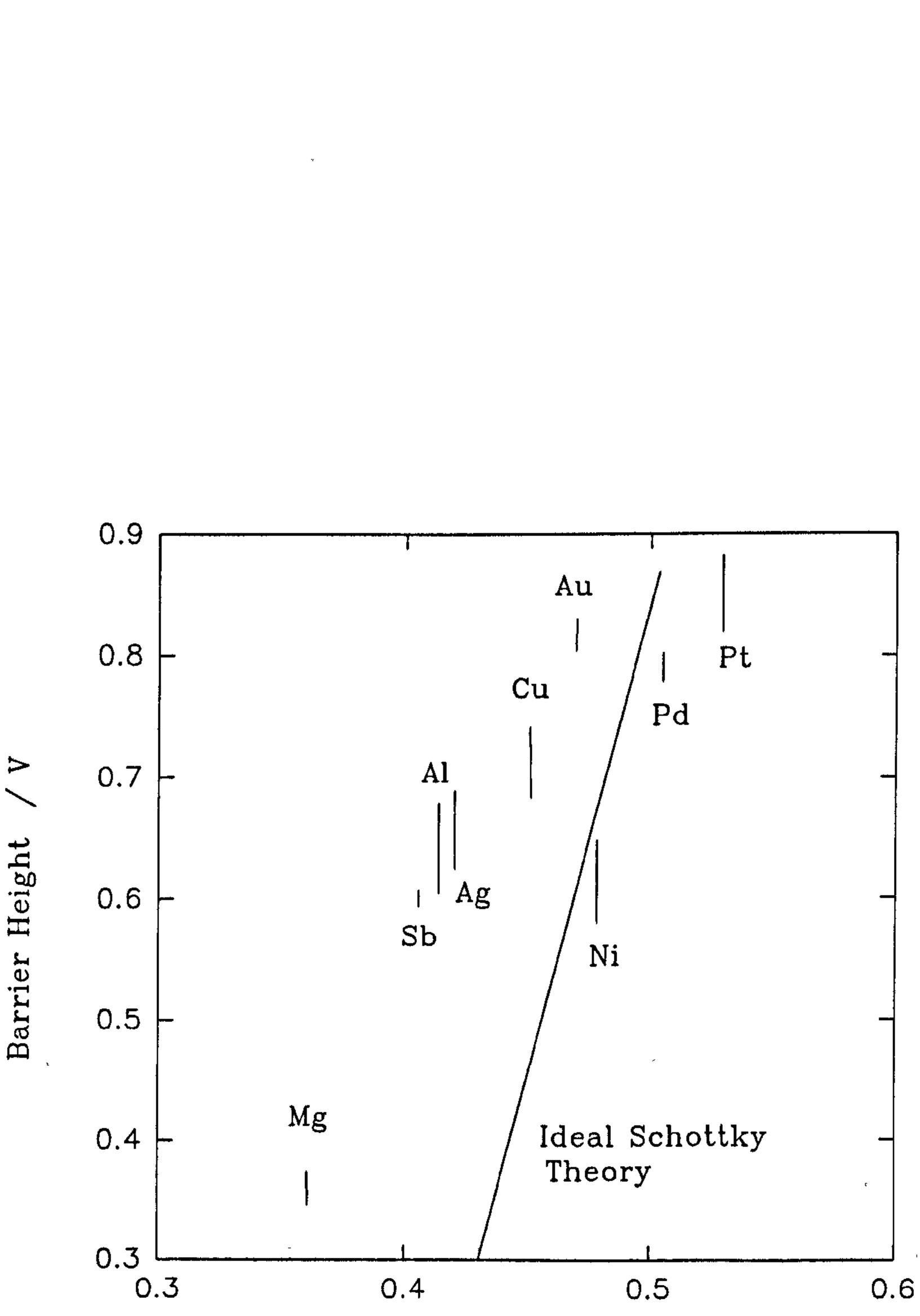


Fig. 3.7 Experimental Barrier Heights to n–type Si. Full Line Represents ideal Schottky theory including image force lowering. (After Cowley and Sze, 1965)

where the symbols have their usual meaning, then the relationship of Rc to ϕ may be summarised thus;

light doping ($kT/E_0 >> 1$)

$$R_c \propto e^{\left(\frac{\phi}{kT}\right)} \qquad \text{thermionic emission} \qquad \textbf{3.3}$$

heavy doping ($kT/E_0 \approx 1$)

$$R_c \propto \exp\left(\frac{\phi}{\coth\left(E_0/kT\right)\sqrt{N_d}}\right) \qquad \text{thermionic field emission} \qquad \textbf{3.4}$$

Very heavy doping or at low T, ($kT/E_0 << 1$)

$$R_c \propto \exp\left(\frac{\phi}{\sqrt{N_d}}\right) \qquad \text{field emission} \qquad \textbf{3.5}$$

Where $kT/E_0 \geq 1$, the potential barrier narrows such that tunnelling is the dominant means of charge transfer, and so the barrier height, which may only alter by a factor of 2 or 3 with different metals, is of little significance. In practice, values of $N > 10^{18}$ cm^{-3} readily allow contacts to be made with Al. To measure at liquid He temperatures, it is useful to have degenerate Si with $N > 10^{19}$ cm^{-3} to prevent contact freeze-out. The $\exp(N^{-1/2})$ dependence of R is observed, in III-V systems, only for Al on GaAs contacts grown in-situ by MBE: other methods of preparation lead to deviations from this theory. These deviations have been explained in terms of a thin amorphous region at the interface, but no clear model of the conduction mechanism has emerged (see for example Dingfen et al 1986).

3.5.2 Passivation

As with all single crystals, the surface of Si must undergo reconstruction. For Si <100>, the reconstruction is from the diamond unit cell. This study was performed wholly on <100> Si for the same reason that it is the industry standard - the surface state density of the $Si0_2$:Si interface is lowest in this orientation. The disruption of periodicity of the lattice induces surface charge with concentration which can be of the order of the number of surface atoms ~ 10^{15} cm^{-2}. Passivation for Si, commonly by oxidation, can reduce this to ~ 10^{11} cm^{-2} - because bonding to O moves surface energy bonds out of the Si bandgap - and to ~10^{10} cm^{-2} after H or HCl annealing at, typically, 400 C. It is necessary to outline the principle of operation for the MOS system for the Zerbst measurement technique which features in Chapter Four. This is also a useful basis for the discussion of results in this chapter, Section 3.5.4, so it is given here, for the readers convenience.

3.5.2.1 The MOS Capacitor

The MOS capacitor is comprehensively reviewed by Nicollian and Brews 1982; thus, this discussion is restricted to a brief introduction to the basic physics pertinent to the discussion of oxide effects in Section 3.6.3 and the Zerbst technique described in Chapter Four.

The ideal MOS capacitor is shown in Fig 3.8, where t_{ox} is the gate oxide thickness and V_g is the applied gate voltage relative to the p-type substrate. The ideal MOS capacitor is defined as one with no work function differences, no surface states and no oxide charge. Furthermore, it is assumed that the gate oxide is ideal i.e. zero conductivity. Fig 3.9 is the ideal high frequency C-V plot. For negative V_g, the capacitance tends to that associated with the gate oxide, C_0: the band diagram is shown

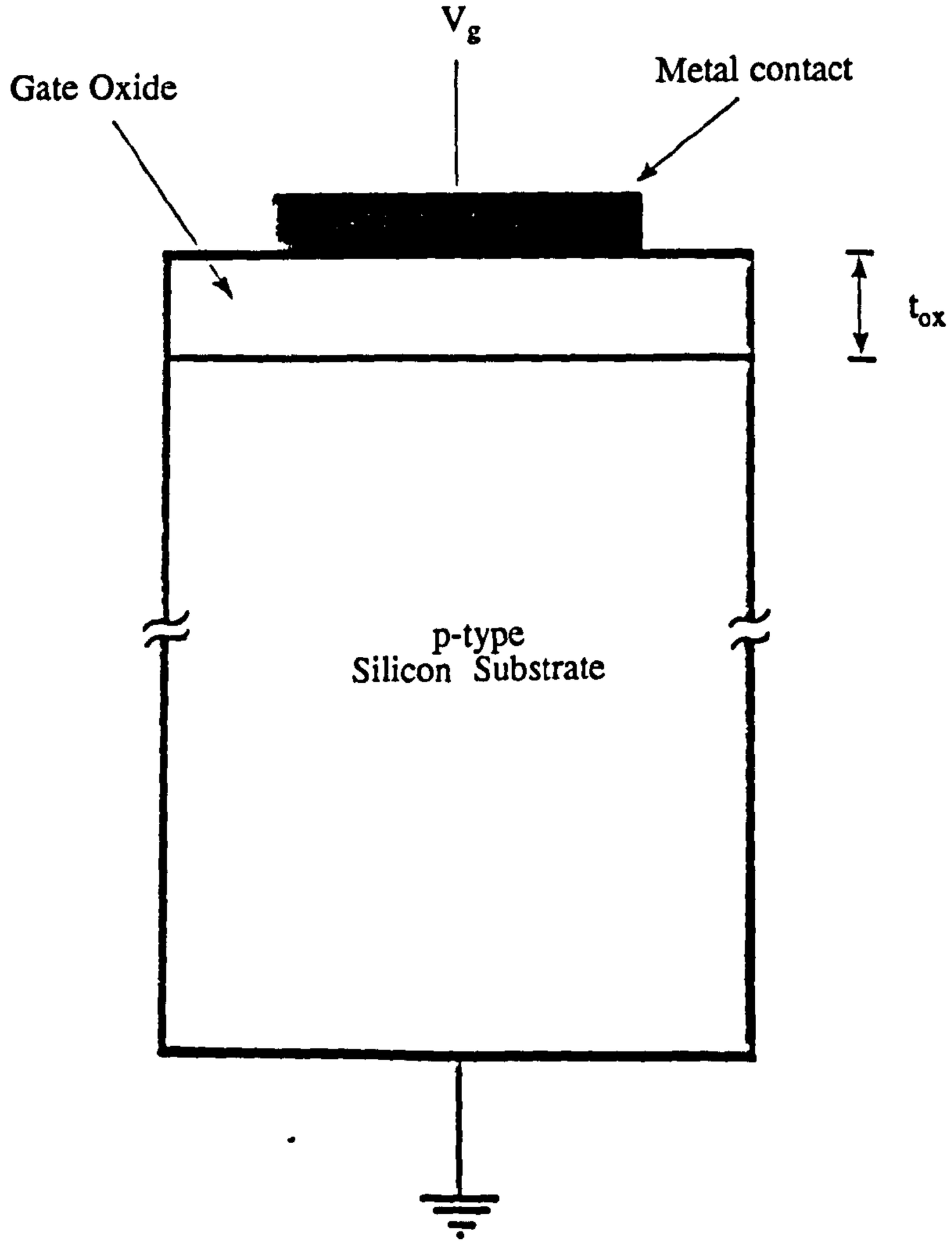


Fig. 3.8 Schematic cross-section of an ideal p-type MOS capacitor.

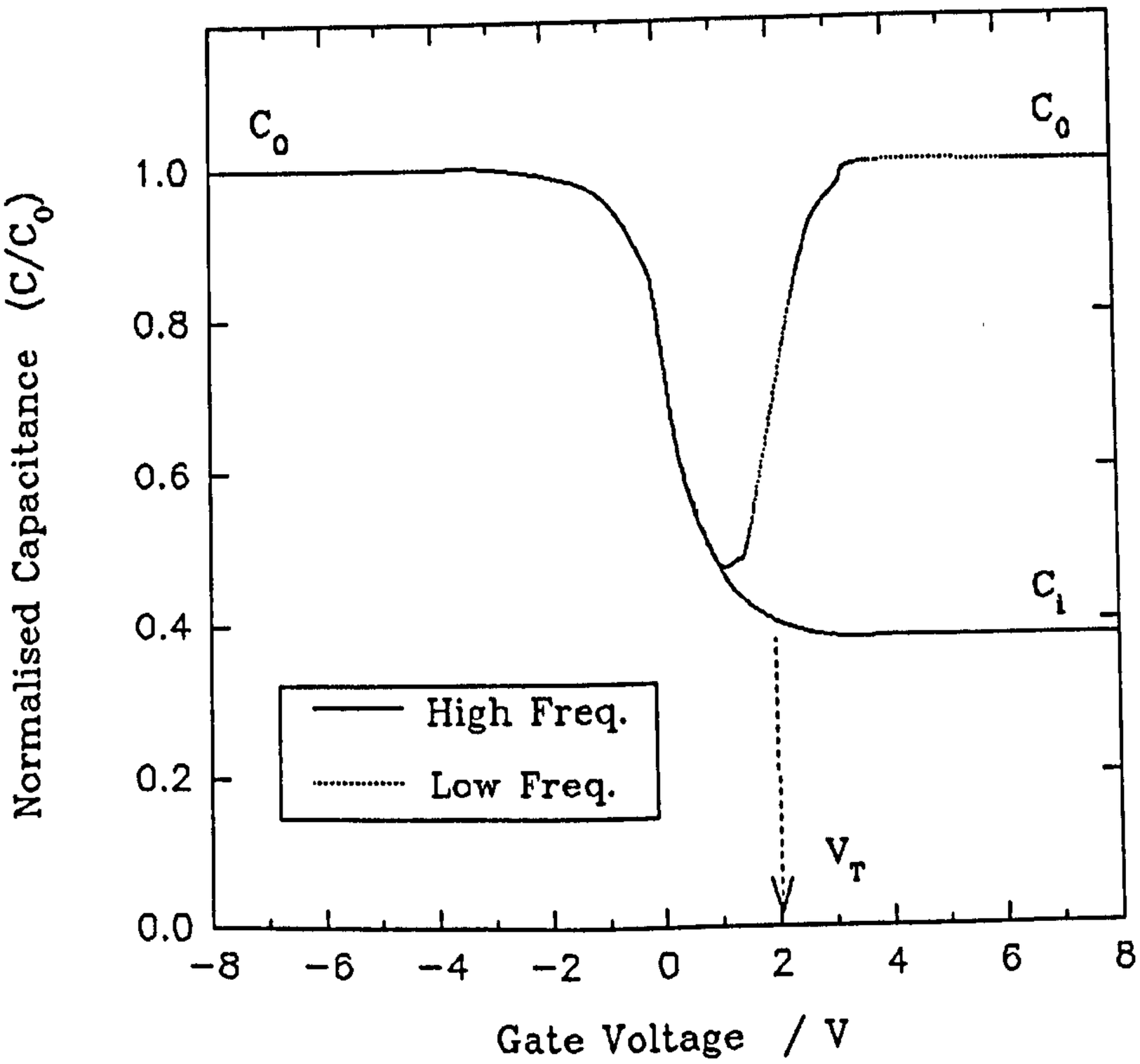


Fig. 3.9 Ideal MOS capacitance–voltage curves for high and low measuring frequencies. (V_T is the threshold voltage, C_i the inversion capacitance, C_0 the oxide capacitance.)

in Fig. 3.10a. The bands at the interface bend upwards as holes accumulate at the Si/oxide interface in response to negative gate charge. Since the accumulation charge Q_{acc} is an exponential function of the surface potential φ_S, the associated accumulation capitance C_{acc}, increases rapidly with increasing V_g.

$$C_{acc} = \frac{dQ_{acc}}{d\psi_s} \rightarrow \infty \qquad \text{for } V_g << 0 \qquad \textbf{3.6}$$

The oxide capacitance is in series with the accumulation capacitance, hence the total capacitance, C_t, tends to C_0.

For V_g positive, the charge on the gate electrode is balanced by depletion charge from the acceptor atoms under the gate oxide. Clearly the total capacitance is C_0 in series with the surface depletion capacitance, hence C_t falls for increasing V_g. This is illustrated by Fig. 3.10b. Using the depletion approximation, Poissons equation is readily solved to express V_g as a function of depletion distance, W;

$$V_g = \frac{qN_a}{\varepsilon_0}\left[\frac{Wt_{ox}}{\varepsilon_{ox}} + \frac{W^2}{2\varepsilon_r}\right] \qquad \textbf{3.7}$$

ε_0 and ε_{0x} are the permittivity associated with vacuum and SiO_2 respectively, ε_r the relative permittivity of Si. Fig 3.10c illustrates the flat band condition, where no potential gradients exist in the system. For the ideal MOS, flatband capacitance occurs at $V_g = 0$ and $C_t < C_0$. This is because C_{acc} is finite and is a function of the substrate doping.

As V_g is made more positive, the conduction band moves through the Fermi level, which is fixed due to the assumption of zero current flow. The surface is therefore inverted i.e. n-type. The inversion charge is exponentially dependent on ψ_s (cf accumulation charge). This has two consequences; firstly, the inversion region remains quasi 2D, i.e contains quantised energy levels and the wavefunction is confined

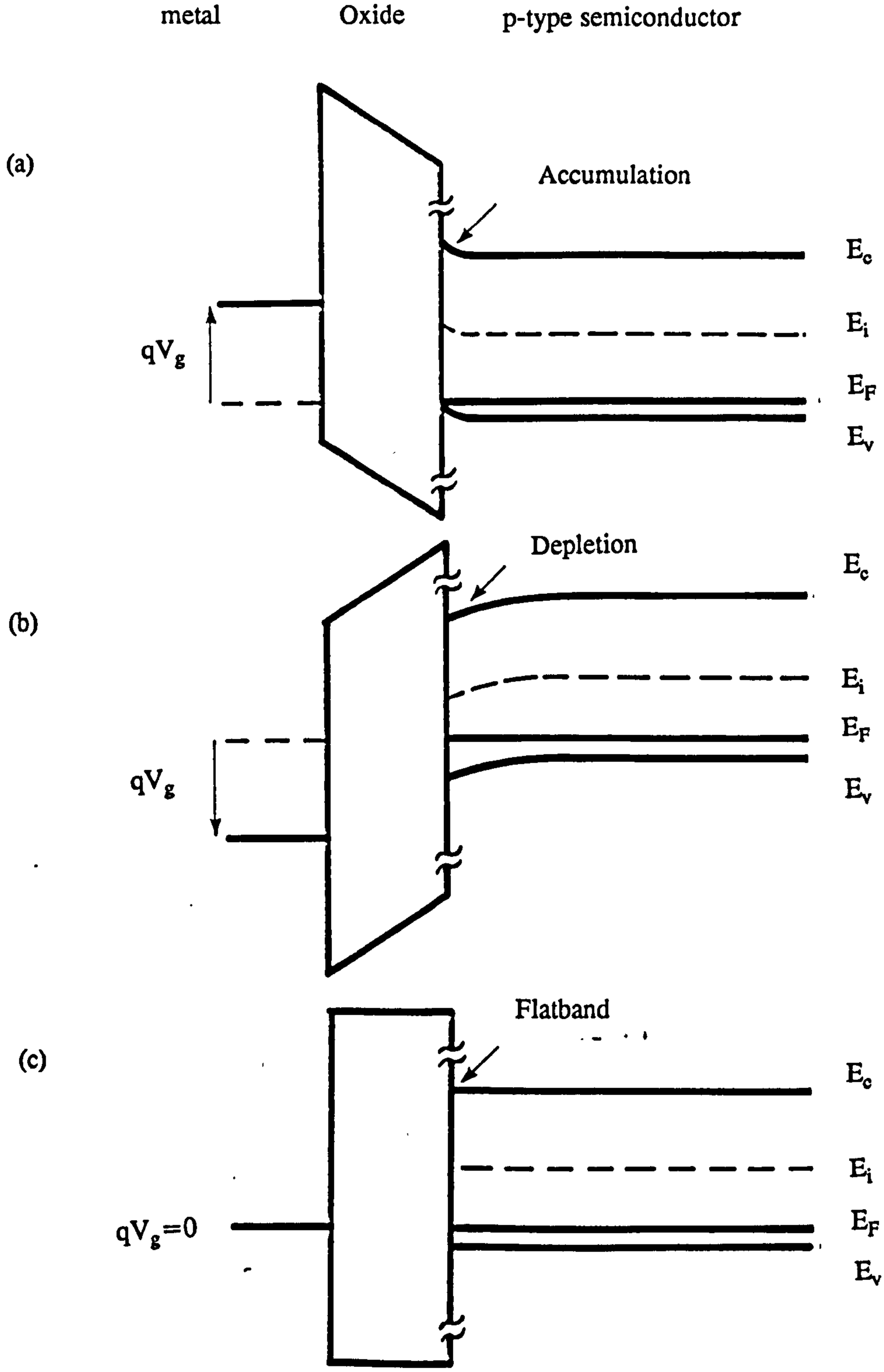


Fig. 3.10 Ideal energy band diagram for a p-type MOS capacitor with (a) corresponding to accumulation, (b) corresponding to depletion and (c) corresponding to flat band.

in the z direction, and secondly, the inversion region screens the depletion region such that C_t remains nearly constant for $V_g >> 0$.

Alternatively, the C-V plot can be obtained at frequencies sufficiently low compared to the electron recombination-generation rate that the minority inversion charge can follow the AC signal. In this case, modulation of charge begins at the inversion layer rather than at the depletion edge. This results in a large increase in capacitance at inversion, directly analagous to the accumulation condition. Thus, C_t tends to C_0, the gate oxide capacitance. For high quality Si this requires modulation frequencies < 100Hz and low frequency C-V plots are obtained via the quasi static measurement technique, Bergland, 1966. This measurement is particularly useful to establish fixed oxide charge concentrations, Ziegler and Klausmann, 1974.

3.5.2.2 Surface States

The various sources for allowed energy levels within the band gap at the $Si:SiO_2$ interface have been summarised in Nicollian and Brews, 1980. To understand the electrostatic effect of surface states, consider an interface characterised by a monoenergetic level at energy E_s (Fig.3.11). When V_g is sufficient to cause the Fermi level to cross E_s, then the charged state of this level will change. Thus, the occupancy of this energy level is voltage dependent, which leads to a distorted C-V curve. For a distribution of energy levels in the band gap, the plot is "smeared". The effect of Q_{it} is to introduce a conductance G, in parallel with the depletion capacitance.

$$G = \frac{C_{surf}\omega^2\tau}{1+\omega^2\tau^2} \qquad \text{with } \tau = C_{surf}.R_{surf} \qquad 3.8$$

Surface states can be donor-like or acceptor-like and can exchange charge either with Si conduction or valence bands or with states in the oxide. This exchange exhibits a characteristic time constant and is broadly categorised as a fast or slow state. Fixed

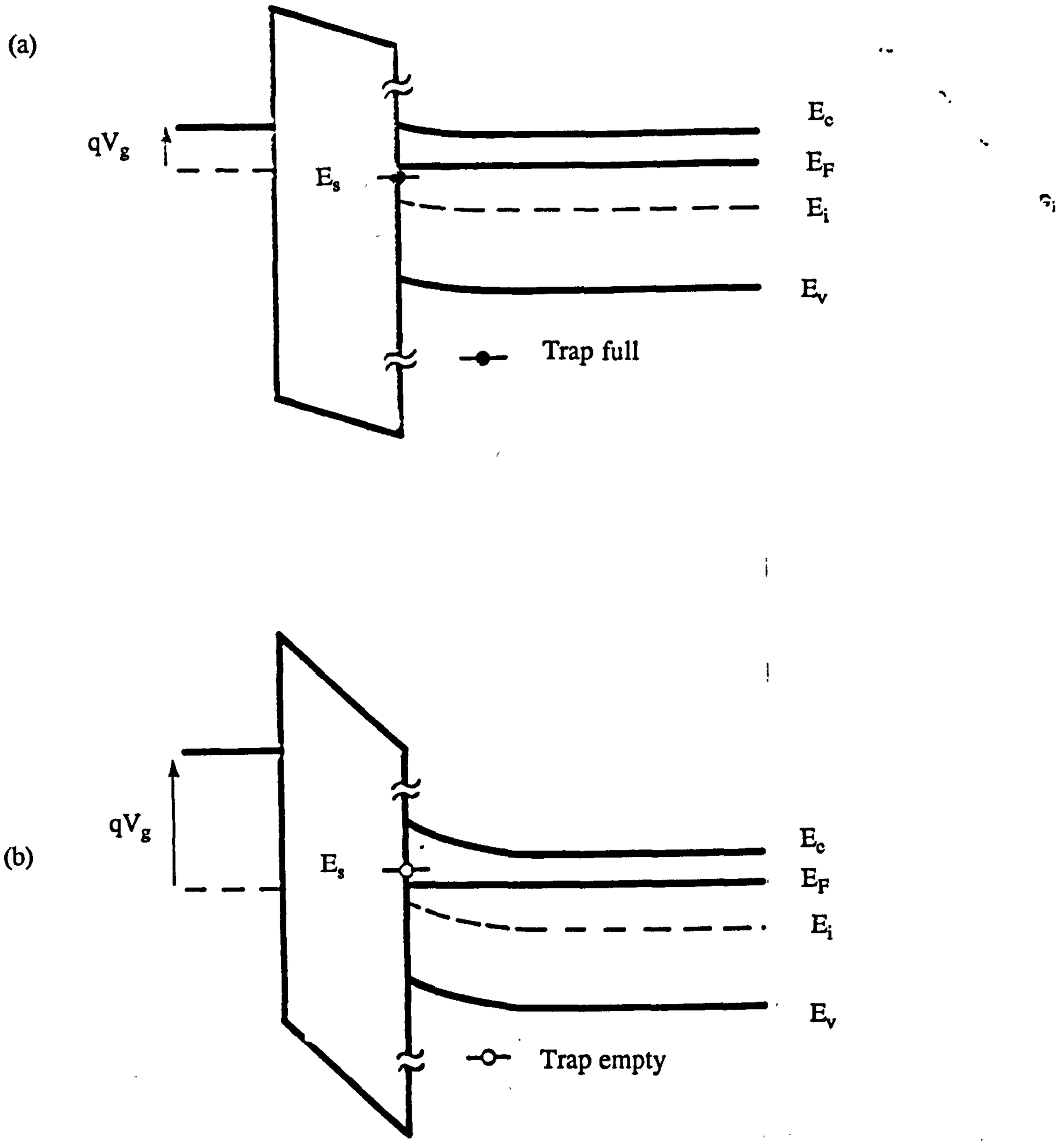


Fig. 3.11 Band diagram showing a midgap state at a Si/SiO_2 interface emptying as a result of an applied gate voltage.

oxide charge due to dangling bonds at the interface are fast states at room temperature and can be reduced to concentrations $\sim 10^{10}$ cm^{-2} by H_2 annealing at low temperature (400 C).

In addition to Q_{it} there exists fixed oxide charge Q_f due to excess trivalent Si and mobile ionic charge Q_m. Q_f is modelled as a sheet of positive charge at distance x_i from the gate contact. This charge will induce an image charge on the gate and in the Si. From Gauss' law, the charge induced on the gate by a charge centroid is:

$$Q_{fg} = \left(\frac{t_{ox} - x_i}{t_{ox}}\right) Q_f \qquad \textbf{3.9}$$

With a non-zero V_g, the total charge on the gate is the superposition of Q_{fg} and that due to the voltage dropped across the gate oxide.

$$Q_g = Q_{fg} + C_{ox}\left(V_g - \psi_s\right) \qquad \textbf{3.10}$$

The gate voltage required to restore flat band, V_{fb}, can be found by integrating for an arbitrary distribution of oxide charge.

$$V_{fb} = \frac{1}{C_{ox}} \int_0^{t_{ox}} \left(\frac{t_{ox} - x_i}{t_{ox}}\right) Q_f(x_i)\, dx_i \qquad \textbf{3.11}$$

Thus Q_f can be measured. Note that positive oxide charge will induce a negative shift in V_{fb} for n or p-type Si, while the magnitude of this shift is directly proportional to the charge:

$$\Delta V_{fb} = \frac{Q}{C_o} \qquad \textbf{3.12}$$

Mobile ionic charge is a result, usually, of alkali metal ions in the oxide, particularly Na. These ions are sufficiently mobile to drift within the oxide when low V_g are applied. Negative V_g causes the positive alkali metal ions to migrate to the gate contact/oxide interface where they do not influence V_{fb}. A positive V_g, in contrast, moves these ions to the Si-SiO_2 interface where they have maximum effect. This leads

to instability in the threshold voltage (very large shifts in C-V curves were observed with time in early MOS systems, which were invariably Na contaminated - 4 or 5 volt drifts over lunch). An extremely useful bibilography of early studies on MOS systems is given by Schlegel 1967.

3.5.2.3 Pyrolytic Oxides

In order to discuss the results of the following Section, it is necessary to consider the nature of the pyrolytic oxide. This oxide is deposited by LPCVD at 400^0C using silane and oxygen:

$$SiH_4 + O_2 \rightarrow SiO_2 + 2H_2 \qquad \textbf{3.15}$$

This, and similar, techniques are fully reviewed by Kern and Rosler 1977. Conventionally, at the EMF, this oxide is used as interlevel insulation and phosphine is added to provide 5% wt P doping via

$$4PH_3 + 5O_2 \rightarrow 2P_2O_5 + 6H_2 \qquad \textbf{3.16}$$

P is used to enable the glass to flow, when subsequently heated at 1100^0C, to provide step coverage.

Although no phosphine was used for this work, high backgrounds are expected in the deposition chamber and it is expected that the pyrolytic oxide may contain relatively high P concentrations depending on immediate reactor history. In addition, this oxide is of low density ~ 2 gmcm^{-3}, poor stoichiometry, is inhomogeneous and subject to high pinhole densities. It can also be expected to contain up to 4 wt % silanol (SiOH) and significant moisture content. Note that this is not state-of-the-art pyrolytic oxide, which can be of much higher quality.

3.6 RESULTS AND DISCUSSION

3.6.1 Ohmic Contacts

There are three problems to be addressed; firstly, to put a low contact resistance (ohmic) metallic contact onto the surface, secondly, to contact to a buried layer and, thirdly, to contact exclusively to selected buried layers.

One simple method of forming ohmic contacts is to form a pressure contact with phosphor bronze; these, however, are unreliable and develop high values of R_c and non-linearity below 150K. For rapid "in-house" probing and for e-CV ohmic contacts, a GaIn eutectic is daubed onto the surface: a gold pad on top prevents punch through from the probe needle. Buried layers are contacted by abrading the contact region with a diamond scribe, which both allows the GaIn to form intimate contact with buried layers and the damage introduced lowers the contact resistance.

In this study, for reliable, low temperature surface contacts, Al was evaporated or Al:Si 1% was sputtered in an Ion Tech sputterer. This was followed by an anneal at 500 C for 15 minutes in dry flowing N_2. The sputtered Al:Si approach is also used at the EMF with a 435 C sinter for 10 minutes in N_2, 5 minutes in N_2:H_2 and 5 minutes in N_2. The purpose of the anneal is not to form an Al:Si alloy but to reduce the oxide between the two, hence the use of a inert ambient and the further use of a reducing gas at the EMF. A distinction is made here between alloying which occurs above the eutectic temperature and annealing.

When possible, the contact doping is grown into the epitaxial device and high resistivity substrates are used to prevent parallel transport. (Note that this is not always consistent with the requirements of the device). A common problem is that posed by selective contacting to buried layers. One in-house answer to this problem was the diffusion of dopants from a spun-on liquid source: Emulsitone Silicafilm was used

successfully to contact to substrates, but necessitated a 900 C 10 minute diffusion. This was considered as a possible solution to the problem of selectively contacting to the doping superlattice. This structure, reviewed by Dohler 1986, was intended to produce high carrier mobilities by means of reduced f-phonon scattering and reduced effective mass carriers in high subbands. In the n-i-p-i structure, it is desirable to contact selectively to either the n or p regions. McPhail et al, 1987, used SIMS to demonstrate the smearing effect of anneals on the super lattice profile. This study showed that a 760 C, 30 minute anneal significantly degraded the structure, thus precluding the use of the spin-on dopant.

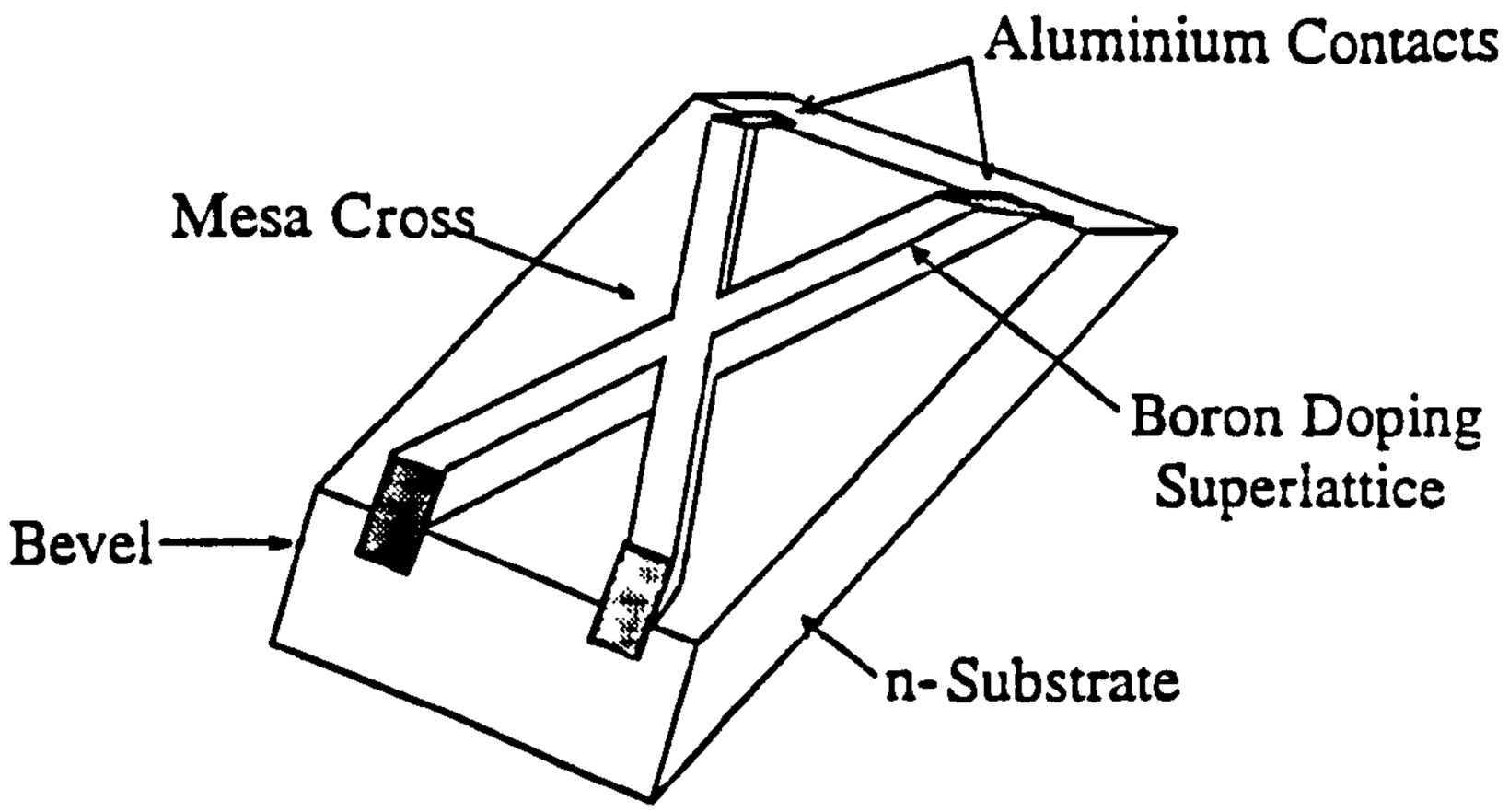
The author considered two solutions. One approach is to alloy Al contacts above the eutectic temperature of 577 C, another to alloy above the melting point of Al, 660 C. These are liquid phase reactions and Al will act as a p-type dopant rectifying to n-type Si: Au-Sb is used as the equivalent n-type contact. In order to allow the Al to reach deep lower layers which may be 0.5 μm below the surface, it is useful to anneal above the Al melting point of 660 C. A deficiency of this approach arises from preferential diffusion into Si defects. This causes spiking with resultant high field regions and can be mitigated against by the use of Si doped Al, which also reduces void formation due to Si precipitation. Hogarth 1986 alloyed contacts at 680 C and 610 C. The higher temperature alloy gave a better morphology but higher R_c: this is in direct contrast to a report by Andrews and Phillips, 1974. The optimum (in-house) solution found for a B doping superlattice (p-i-p-i) is a 680 C, 4 minute anneal combined with a bevel- nominally 5 minute angle - to reveal buried layers. The bevel is achieved by a mechano-chemical method using a alkaline etchant, Syton, with a modified Kent 3 automated polisher. This technique minimises the mechanical damage to regions to which the Al must rectify. The bevel was measured as 32 minutes using a Talystep, which corresponds to a magnification factor of -100. The bevel was then stained using an electrolyte of copper salt in a selenous acid. Under strong diffused white light,

electron hole pairs are liberated producing Cu^{2+} ions, and these preferentially plate n-type material. Al contacts were then evaporated through stainless steel foil masks to cover the revealed superlattice and annealed, Fig. 3.12 (a). The room temperature current voltage characteristics of this contact, to both the p-type delta layer and the n-type substrate, are shown in Fig. 3.12 (b), rectification to the substrate improved at low temperatures, as expected.

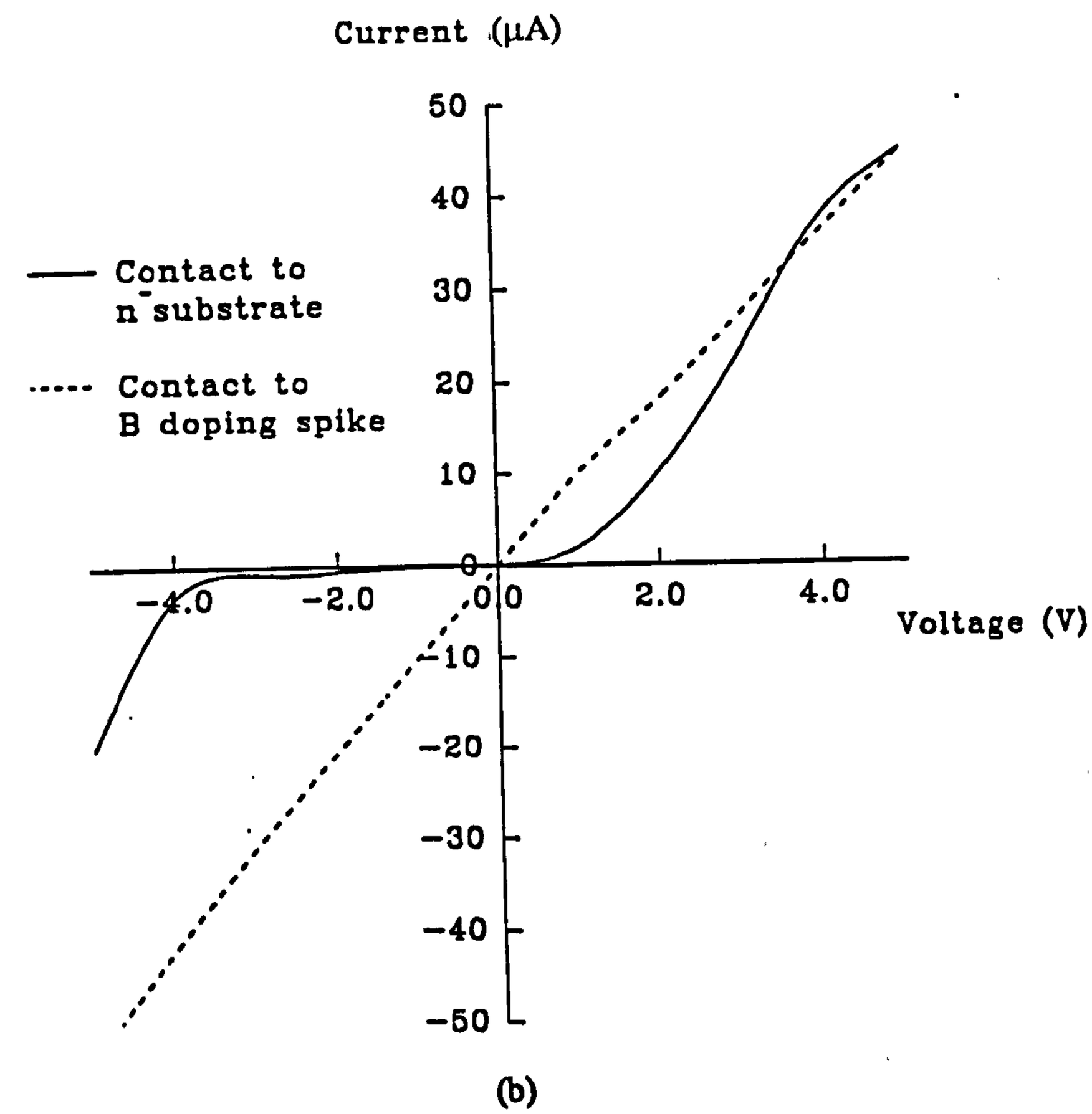
For studies on SiGe remote doped structures or delta layers, the requirement is less severe. The conducting channel is of one polarity only and reliable measurements are often required only for $T < 50K$. The structures are designed such that the background doping is lower than the metal-insulator transition (5×10^{18} for B, 10^{18} for Sb) so that these layers freeze out, i.e free carriers are localised at the parent ions. The conducting buried layer is located within 150 nm of the surface either to allow modulation from a surface gate or to minimise dopant diffusion during subsequent growth (after the channel has been formed). For these structures, an Al contact is annealed at 600 C for 3 minutes to form a eutectic contact to the channel. This contact is adequate for n or p-type structures for sheet concentrations $> 5 \times 10^{10}$ cm^{-2}.

3.6.2 Ion Implantation and Activation

VLSI processed structures are contacted via ion implantation. Multiple ion implants of varying incident energy result in nearly constant doping density throughout the depth of the contact region. The implantation amorphises the contact region. This damage is conventionally annealed out in a tube furnace at, typically, 1000^0C for 60 minutes. Such anneals are not feasible for MBE grown structures. In order to preserve abrupt dopant profiles or to avoid relaxation in strained structures, it is necessary to devise an alternative method of implant activation. At the EMF, simulations of sheet



(a)



(b)

Fig. 3.12 (a) Schematic of the mesa Hall cross used to assess a p-i-p-i superlattice. (b) The room temperature current-voltage characteristics of the contacts to the B doping spikes and n⁻ substrate.

resistance as a function of anneal temperature and time from Suprem 3 were fed into an RS/1 statistical model to solve for concentration, C:

$$\frac{\partial C}{\partial t} = \frac{\partial J}{\partial x} \quad \text{, where} \quad J = -D_{n,p}\frac{\partial C}{\partial x} + \frac{qD_{n,p}}{kT}EC_a \qquad \textbf{3.17}$$

The results of this analysis suggest that, for RTP, the sheet resistance is insensitive to anneal time for short anneal times. From Fig 3.13, it is seen that a 600^{0}C anneal for 10 to 20 seconds results in acceptable sheet resistances of order 100 Ω□ and that this value is constant up to 100nm from the surface. The diffusion coefficient for a given species, D is a function of a number of parameters. For B, D is given by

$$D = \left\{\left(D^x + D^+\frac{n_i}{n} + D^-\left(\frac{n}{n_i}\right) + D^{--}\left(\frac{n}{n_i}\right)^2\right)F_p + D_{oed}\right\}F_{ted} \qquad \textbf{3.18}$$

where D^x is equal to neutral diffusion, D^+ is positive charge state diffusion, D^- negative charge state diffusion, D^{--} double negatively charged state diffusion, F_p an enhancement factor due to high P concentrations and D_{OED} an enhancement due to oxidation. Each charge state diffusion is of the form

$$D^n = \text{Constant}.e^{\left(\frac{-E_n}{kT}\right)}$$

For B, D was found most sensitive to D^+. For relatively long anneals (>60s), the sheet resistance is 200 Ω□, for D^+=2.7, and 100 Ω□, for D^+=4.0. For anneals less than 20 seconds, however, the sheet resistances are comparable for either value of D^+.

Unfortunately, the RTP equipment at the EMF was inoperable for much of this study. Van Gorkum has used a long, low temperature (600 C) anneal to activate implants into MBE grown Si. However, the author found that a 15 second anneal in a tube furnace at 1000 C in an N_2 ambient provided 80% implant activation, as measured from sheet resistances. The temperature of the sample during this anneal is unknown. Electrochemical CV measurements by the author (using a Polaron 4200) of B doping

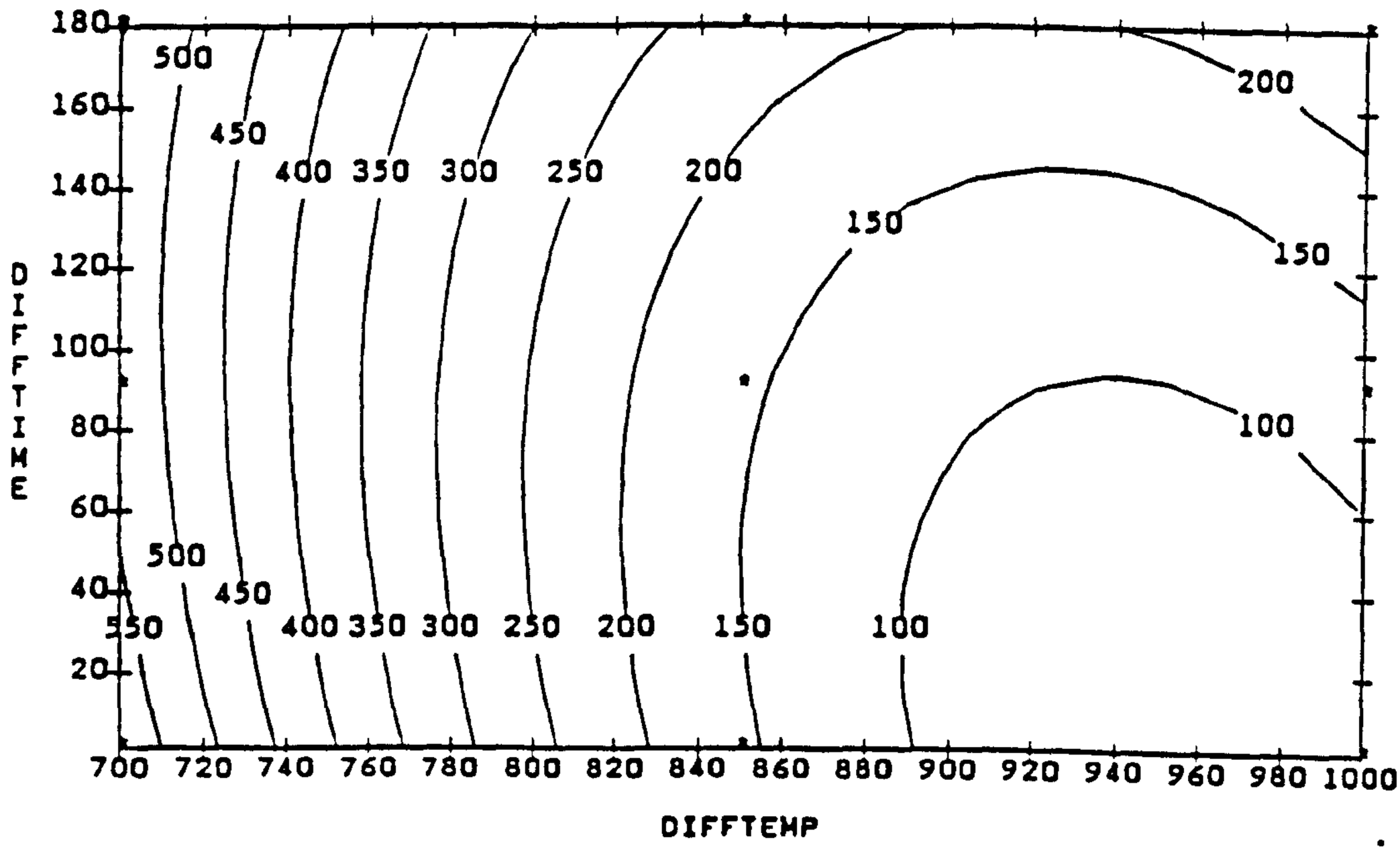


Fig 3.13 Sheet resistance ($\Omega\square$) variations with diffusion time and temperatures for a dual B implant of $6x10^{15}$ cm^{-3} at 90keV and $2x10^{15}$ cm^{-3} at 25keV.

staircases grown by MBE detected no evidence of dopant diffusion for this anneal (or a 900 C 30 sec anneal that caused ~45% activation). The depth resolution of these measurements was, like all CV profiling techniques, limited by the Debye length (see for example Johnson and Panousis, 1971), so a further test was performed to assess this activation process by Biswas, 1992. This SIMS analysis of delta layers in Si (Fig 3.14) shows that the anneal has broadened the delta by a maximum of one nm so the FWHM remains narrower than the ground state wavefunction.

3.6.3 Schottky Contacts

Schottky metal/semiconductor contacts (of which the 'ohmic' contact is a special case) provide a means of space charge modulation without the use of a dielectric. From the analysis of theory in Section 3.5.1 it is clear that such contacts will not have good rectifying properties when made to high doped Si ($>10^{17}$ cm^{-3} for Si). In addition, it is clear that surface preparation is a more important parameter than the metal used. Schottky contacts manufactured on MBE Si throughout the majority of this study yielded poor rectifying properties (e.g. $<\sim$3V breakdown) irrespective of pre-clean or the metal used e.g. Pt, Ti, Al, Au:Sb. However, contacts to substrate material processed simultaneously with the MBE Si as controls often provided good diodes. It is suggested that this was the result of contamination in MBE Si.

Modifications to the V90S, which are strongly believed to result in reduced metal (Cu) contamination, (fully described in Chapter Five) were undertaken during 2D SiGe studies. Schottky contacts were made to three Si p^- layers that were grown consecutively after these modifications with T_s the variable parameter. The generation lifetimes were measured in these layers as part of a Zerbst study, (see Chapter Four, pages 78-82). The I-V characteristics for Schottky contacts made to the two 'good' layers ($\tau_g \approx 1$ μs) and the 'poor' layer ($\tau_g \approx 1$ ns) are shown in Fig. 3.15. This evidence does not contradict the belief that the electrical characteristics of Schottky contacts

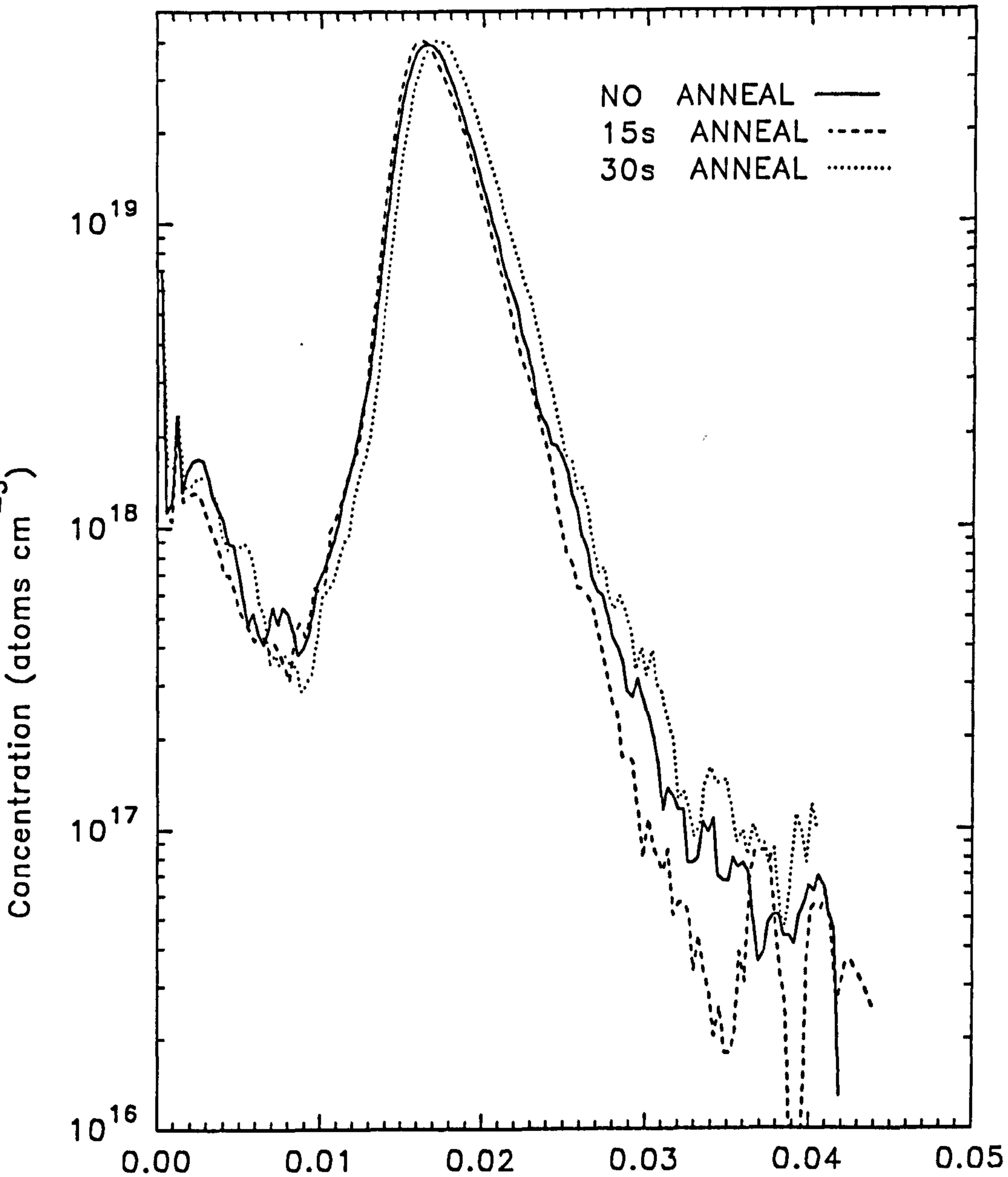


Fig. 3.14 SIMS profile of a B delta layer annealed at 975 C for 15 and 30 seconds. (Profiling conditions; 2keV O_2^+ at normal incidence.)

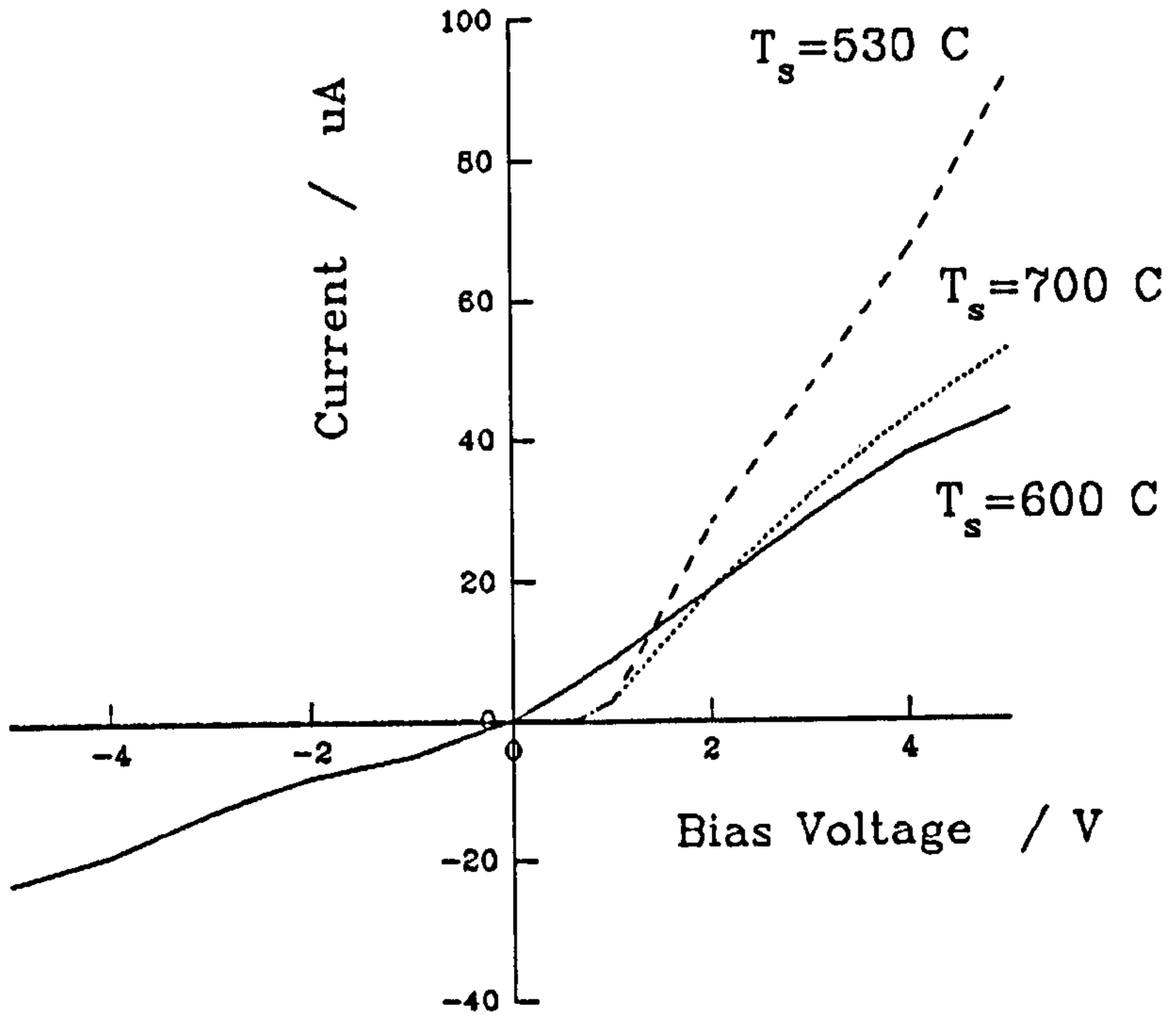


Fig. 3.15 Current-voltage characteristics of Ti/Al Schottky diodes formed to MBE Si. The analysis of the generation lifetimes found in these layer is discussed in Chapter Four.

made to MBE Si were determined by material quality. These sputtered contacts consist of 1 μm Ti capped by 0.5 μm Al on Si that had been pre-cleaned using the RCA substrate clean as detailed in Section 2.2. The Al is necessary to allow gold wire bonds to be made to the contacts and is in-situ sputtered to prevent Ti oxidation which otherwise results in poor Ti:Al adhesion. No anneal is necessary.

3.6.4 Substrate Cleaning

It was noted in Section 2.2 that 4" substrates were given no ex-situ clean, but that 3" were given full RCA cleans. The difference between 4" and 3" ex-situ cleans for the V90S is due to a substrate holder design that leads to temperature gradients across 3" wafers during the in-situ clean resulting in slip lines. The 3" wafers are necessary for layers requiring processing at the VLSI facility accessed, so the author developed the strategy of an RCA clean (which leaves only a very thin oxide on the Si that can be desorbed at a lower temperature, 800 C). This clean involves the stripping of native SiO_2 by 25% HF (2 mins), an organic contamination clean and thin (~3 nm) SiO_2 growth in hot ammonium hydroxide and H_2O_2 (20 mins), further oxide strip in 5% HF (1 min), metal clean and thin SiO_2 growth in hot HCl and H_2O_2 (20 mins), followed by a cascade rinse in 18MΩ H_2O. It was found that dislocation levels in epilayers grown on substrates given this treatment (~10^5 cm^{-2}) were higher than those subject only to an 800 C in-situ clean, unless VLSI Selectipur grade chemicals were used in a class 10,000 fume cupboard, in which case they were at the levels generally achieved by the 900 C in-situ clean alone, Table 3.1. Therefore, the hypothesis that dislocation concentrations were dominated solely by residual SiO_2 appears improbable, further underlining that the mechanisms of defect formation in MBE Si are poorly understood. Note that dislocation concentrations have since fallen in Si grown in the V90S (see Table 4.1): any study of the effects of growth technology on defects requires

Ex-situ Clean	In-situ clean	Mean s-pits (bulk) $x10^3$ / cm^{-2}	Mean s-pits (interface) $x10^3$ / cm^{-2}	Mean Dislocations $x10^3/cm^{-2}$
None	800 C	50 ±40	80 ±60	400 ±300
None	900 C	7 ±3	7 ±5	0.2 ±0.1
RCA (Analar)	800 C	40 ±2	200 ±100	3,000 ±1,000
RCA (Merck, VLSI grade)	800 C	8 ±4	3 ±2	0.3 ±0.1

Table 3.1 Assessment of substrate clean procedures for 3" substrates in the V90S. Shown are defects etch counts from B doped epilayers (~0.75 μm), grown on Sb doped Monsanto 3" substrates, 0.02-0.001 Ωcm.

lengthy compilation of results and statistical analysis, this is currently in progress by other workers for the V90S, but is not yet complete.

3.6.5 Surface Currents

In this Section the effect of surface leakage currents on device measurements is discussed. For low voltage measurements such as the Hall effect, leakage currents have been found to be unimportant, but for reverse current bipolar measurements these currents can be greater than the bulk current. The author's early measurements on mesa pn diode structures grown by MBE, but without passivation, yielded poor reverse characteristics (e.g. leakage currents $>$ ~10 μA at 1V reverse bias, 1 mA at 3V). This is contrast to work by Bean 1984 who reported picoamp reverse currents at 10 V without passivation, and it is possible that the high leakage currents observed represent a surface effect due to the nature of the surface after etching or some other subsequent contamination. Accordingly, further pn diodes where grown and processed, at the EMF, using mask set Eμ 726. Details of growth and material quality assessment from Smith et al 1990 are given in Chapter Four, while the structure is shown in Fig3.16.

The basic theory of pn junction diode operation is detailed in most semiconductor text books or indeed, in Shockley's orginal paper of 1949, so the results will be quoted without proof.

The reverse current of a p^+ n diode is governed to a first approximation by

$$J_r = q\sqrt{\frac{D_p}{\tau_p}\frac{n_i^2}{N_d}} + \frac{qWn_i}{\tau_0} \qquad \text{with} \quad \tau_0 = \frac{\tau_n + \tau_p}{2} \qquad 3.18$$

The first term is the diffusion current, $J_{d,}$ arising in the neutral bulk, the second the current, $J_{g,}$ due to generation of carriers in the depletion region. J_d is a constant and, in Si for T $<$ 300K, will be negligible compared to J_g.

$J_g \propto W \propto V^{\frac{1}{2}}$ for the case of the abrupt junction.

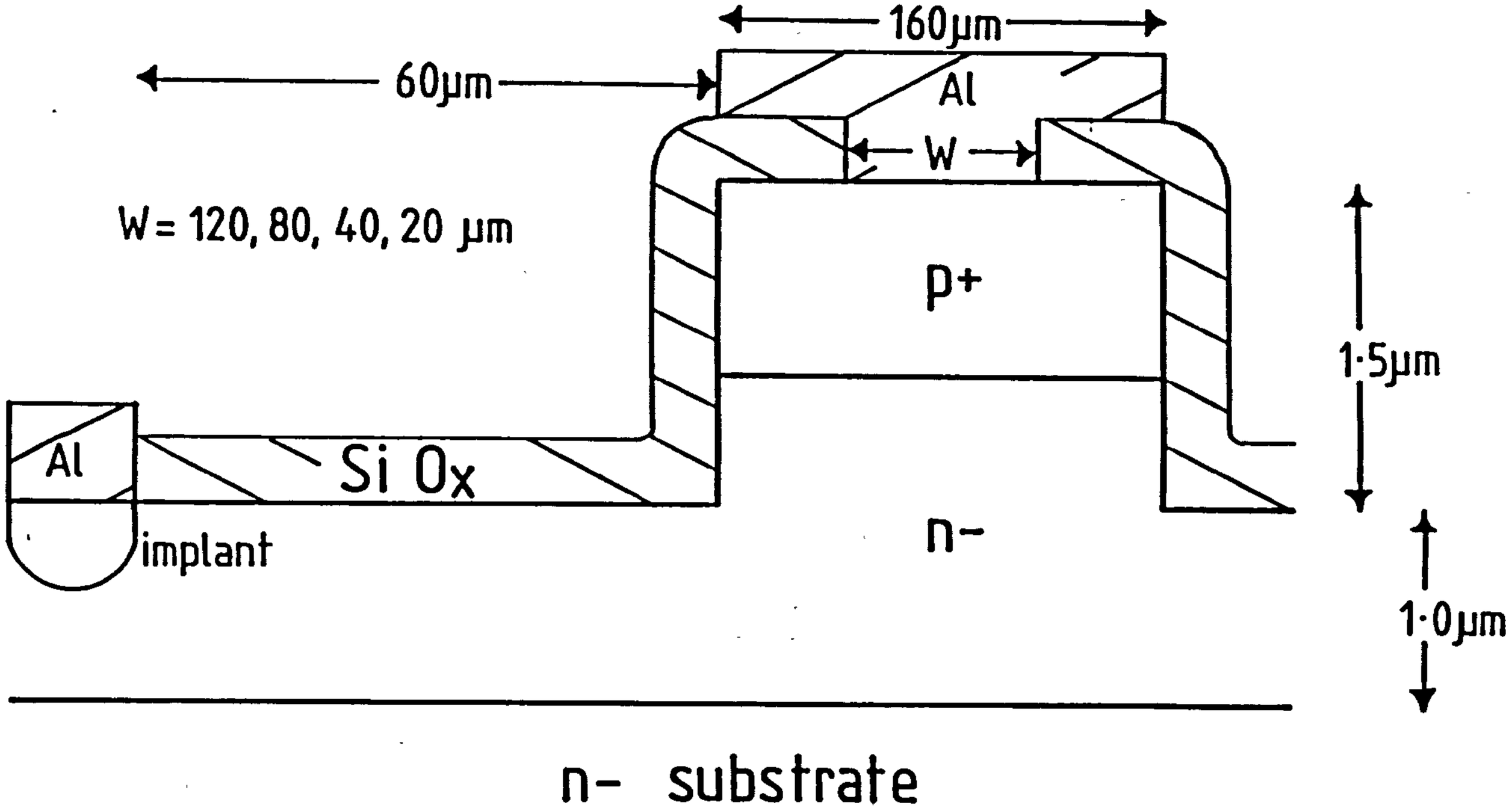


Fig 3.16 Cross section schematic of the processed MBE grown p^+n diode using a pyrolytic oxide as passivation.

Fig 3.17 (a) shows typical forward and reverse characteristics. The breakdown voltage, V_b, of 70V, is in accordance with theory for an abrupt junction and its temperature coefficient indicated that avalance breakdown is occuring. Importantly, for the discussion following, the voltage drop must occur predominantly across the depletion region. Of ~ 1000 diodes on each of two wafers, some 90% yielded rectifying properties: V_b varied by ± 5V as defined at I_r = 100 μA. The observed 'hard' breakdown characteristics appear to indicate that the density of metallic precipitates is not significant; further, that the effect of any high field regions at contact or mesa corners are small. However, this observation is not a strong indicator due to the effects discussed below.

There are two distinct regions of operation, V_r < 18V and V_r > 18V. The behaviour for V_r > 18V will be discussed later. Reverse bias C-V measurements exhibit excellent $1/C^2$ dependence (Fig 3.18) as theory predicts for depletion capacitance and are frequency independent - as expected - in the range 10 kHz to 10 MHz. The carrier concentration of 2×10^{16} cm^{-3} inferred from this $1/C^2$ plot in the usual manner, agrees well with that obtained from electrochemical capacitance voltage (e-CV) measurements by the author, Fig 3.19.

<u>$V_R < 18v$</u>

In this region, J_r is of the order 10^{-8} A cm^{-2}. However, J_r does not exhibit good $V^{1/2}$ dependence (Fig. 3.17 (b) - note that the noise limit was ~5×10^{-13} A), and is nearly temperature independent between 77 and 300K. This indicates that J_r is dominated by conduction through the pyrolytic SiO_2. Two possible conduction mechanisms in insulators give rise to temperature independent currents in this temperature range. Tunnel emission is the process of electron tunnelling from the

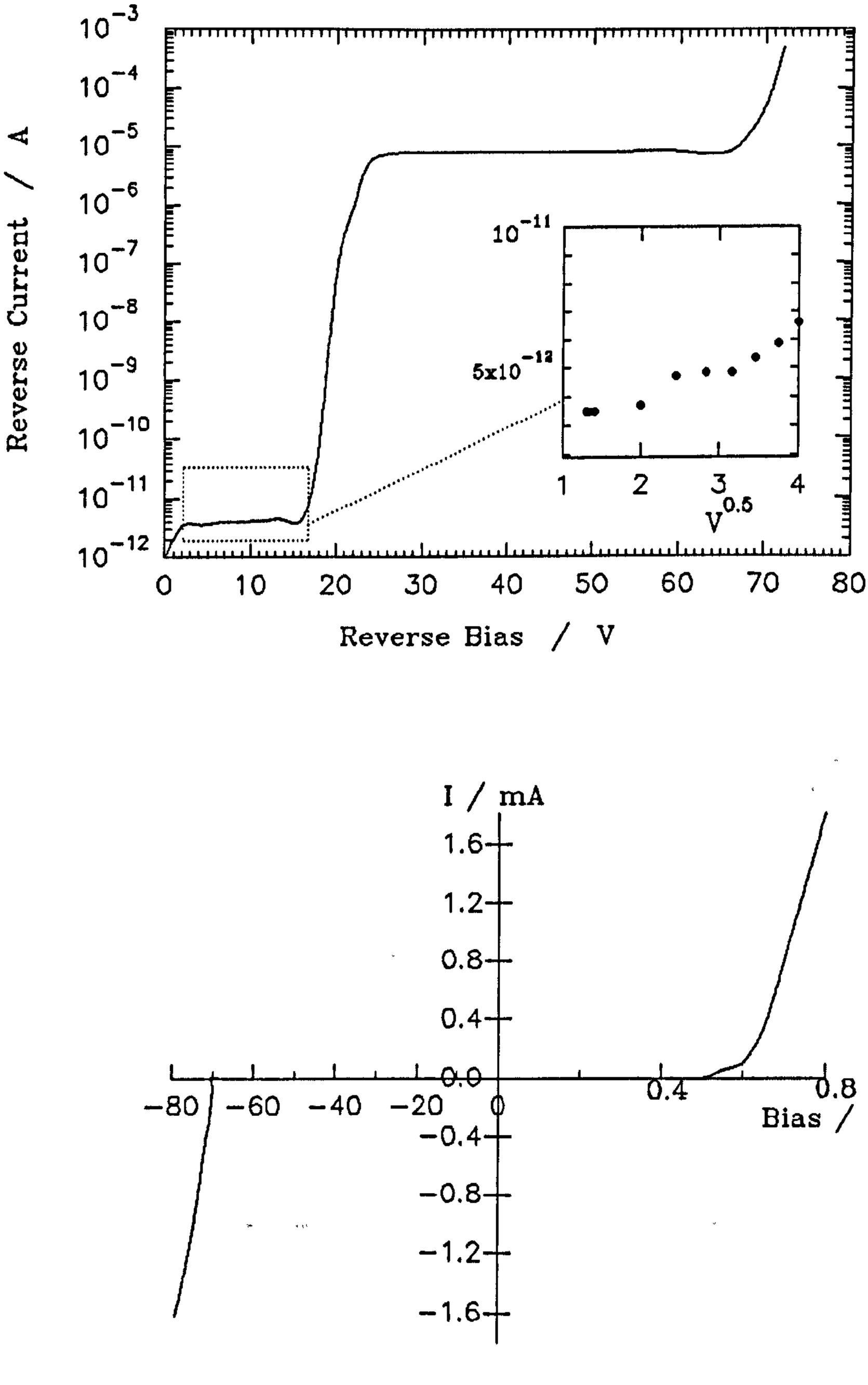


Fig. 3.17 I–V characteristics of p^+n diodes grown by MBE (a) is the linear plot. (b) shows detail of the reverse bias behaviour.

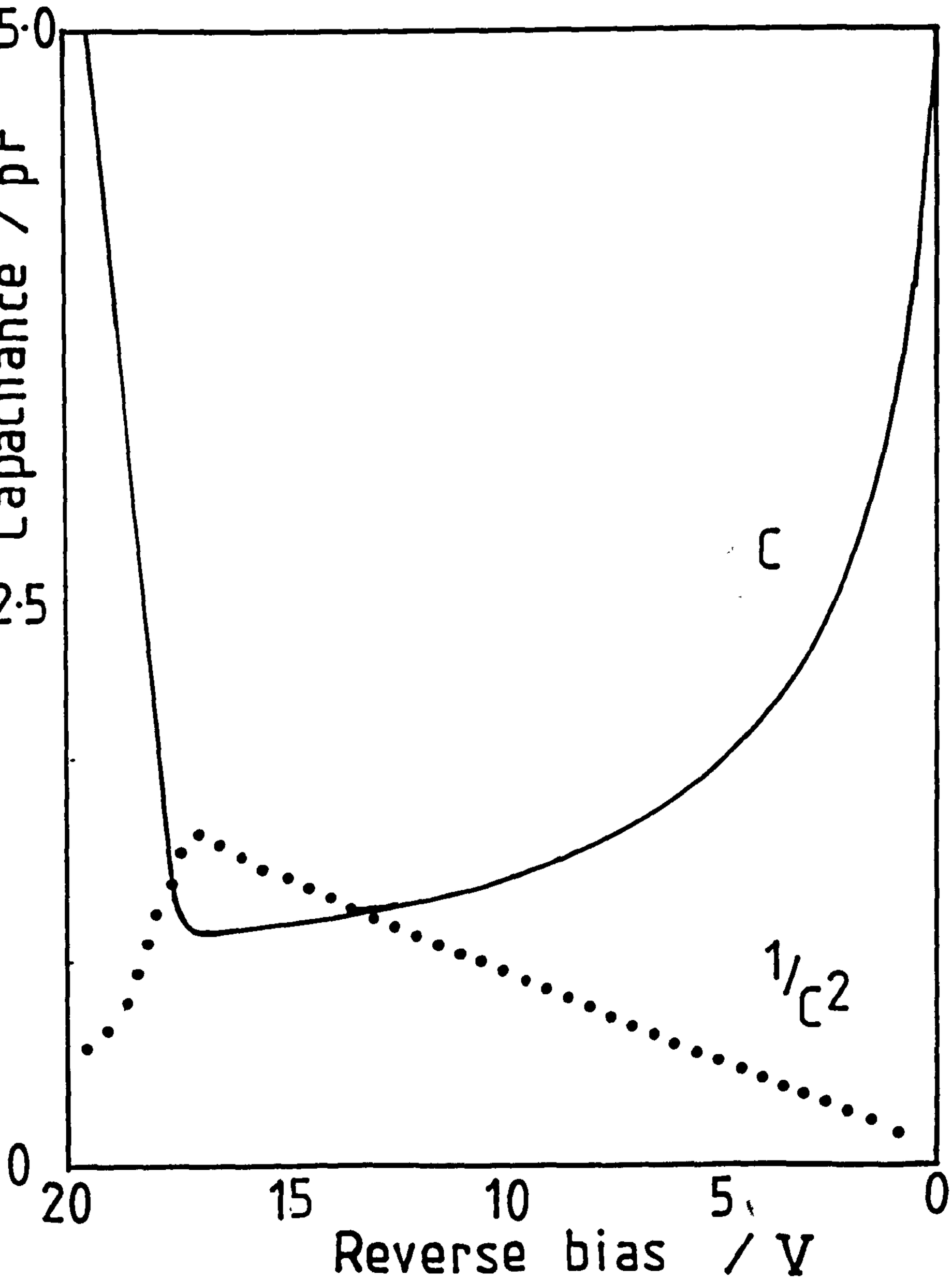


Fig 3.18 Reverse bias depletion capacitance of the MBE grown p+n diode. Note the large rise in capacitance at ~18V.

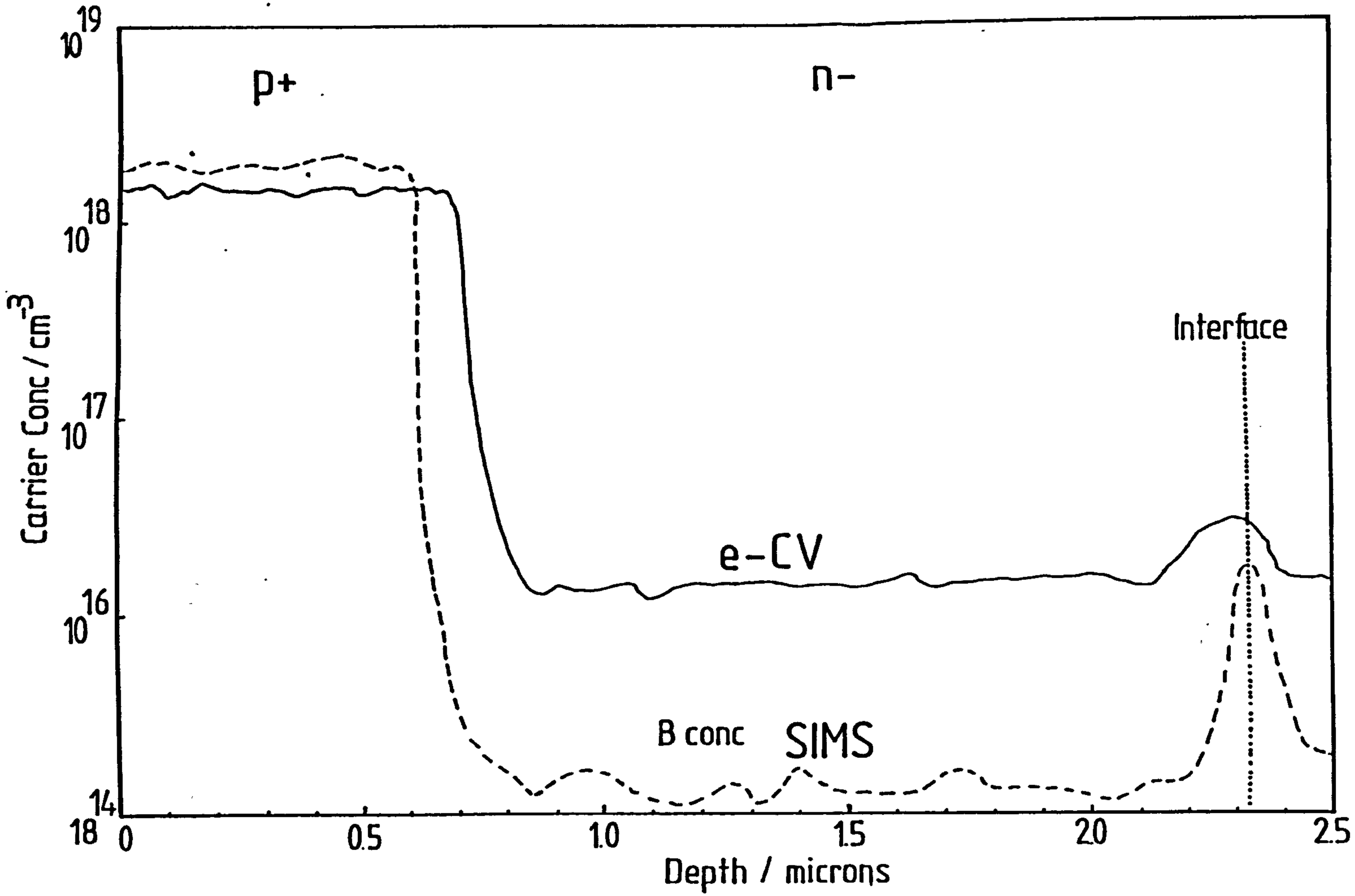


Fig. 3.19 Electrochemical C-V and SIMS profiles of the MBE grown p^+n diode. (Note that the B concentration in the n region is below the SIMS detection limit.)

metal to the insulator conduction band and yields a $J \propto V^2 e^{-\left(\frac{constant}{V}\right)}$ dependence. Space charge limited current is the injection of charge into the insulator where it is not compensated, which leads to a $J \propto V^2$ dependence. Neither dependence can be confirmed by fitting to the data. It is even possible that J_r is of the form;

$J_r = XV^{1/2} + Y V^2 + J_d$, where $Y = A + Be^{-(constant/V)}$, X, A and B are constants. Without an a priori knowledge of the constants, further analysis is intractable. It is possible that the P content of this oxide is significant in providing a source of charge.

$V_R \geq 18V$

A six order of magnitude increase in I_r occurs between 18 and 23V, whereupon the current is approximately a constant until breakdown occurs at 70V. This phenomenon was checked using three different measurement systems to ensure that it was not due to loading effects e.g. range changing.

It can be postulated that this behaviour is explicable either in terms of the formation of a surface channel (at $V_r = 18V$) with saturated current flow, or in terms of conduction through a highly P doped oxide.

Postulate 1 - Oxide conduction.

Conduction through the oxide, either by drifting P^+ ions or via electrons liberated from P or P complexes, seems improbable from a consideration of the current density. The oxide cross sectional area is $\sim 10^{-10}$ m^{-2} , ($J \approx 10^2$ Am^{-2}), which would imply a mobile ion density of 10^{21} m^{-3} , assuming singly charged carriers and a charge velocity of $1ms^{-1}$, although for doubly charged ions the density is more feasible. The onset of current at 20V might be justified by invoking an ionisation potential or an activation energy for P complexes, but it is less easy to postulate a mechanism for saturated current flow. This is not entirely satisfactory.

Postulate 2 - Surface Channels.

The formation of a surface channel could occur in two ways, labelled (a) and (b) below. Both options would require the oxide on the sidewalls to be gated: the upper Al electrode clearly forms a gate on the oxide over the p^+ region surface.

a) The surface of the n^- region could invert to form a continuous p type layer between the two electrodes. Note that $V_g < 0$ will tend to cause accumulation at the p^+ surface. We can estimate the minimum negative oxide charge necessary to induce a 20V shift in the threshold voltage, V_T.. The surface area between electrodes is 2×10^{-8} m^{-2} and t_{ox} is 0.5 μm, thus we would expect C_0 of ~1pF. Assuming uniform distribution of charge in the oxide, this implies $Q \approx 2 \times 10^{-11}$ C, which in turn implies a density of (singly charged) ions $\approx 10^8$ m^{-3}. This is possible.

b) V_T is proportional to N_A which is 2×10^{18} cm^{-3} in the p^+ region. Goetzberger 1966 has computed values of V_T as a function of N_A and t_{ox} for the ideal MOS diode, from which a value of $V_T \approx 20V$ is seen to be feasible. Thus an n type inversion layer forms in the p^+ region, creating a continuous n channel in parallel with the pn diode. The saturation of current through such a layer can be seen as analogous to pinch-off in a MOS FET; in this case, the pinch-off is due to depletion from the grown pn junction. (Analogies can also be made with the gated diode).

Further clues to this mechanism come from the C-V characteristics. The rise in C at 18 V implies the formation of a large capacitance in parallel with the depletion capacitance. This could result from the formation of the inversion layer, but is not likely for the case of an oxide breakdown. Thus it is tentatively suggested that the latter hypothesis - inversion at the surface of the p+ region - is the more likely, although considerable doubt remains.

Whilst further experimentation would be possible i.e. SIMS of the oxide for P concentration, removal of the oxide by etching and device probing (although previous experience of unpassivated probing was contradictive, and the Al would be removed by

the etching), it is important to note that this phenomenon is of interest for this study only in so far as it obscures the nature of the MBE Si diode. In addition, the geometries available on this 'chip' are not suitable for oxide studies. The difficulties encountered led to modifications of the mask design to avoid the problems of gating the passivation oxide. These changes in masks Eμ842 and Eμ931 were successful, as described in Chapter Four. The pyrolytic oxide remained useful as the barrier for all other stages of processing e.g. implantation, mesa etching. Clearly, however, a high quality final oxide was required.

3.6.6 Low Temperature Oxides

A low temperature dielectric is an essential ally of low temperature growth techniques. One attempt to achieve a low temperature oxide was to grow by wet oxidation at 600 C for twelve hours. The 6nm oxide resulting was then used as a seed layer for pyrolytic oxide deposition. This oxide, however, was found to have D_{it} ~ 8 x 10^{12} cm^{-2}, so the acceptability of the long anneal was not assessed. Fortunately, high quality low temperature oxides became available during this study.

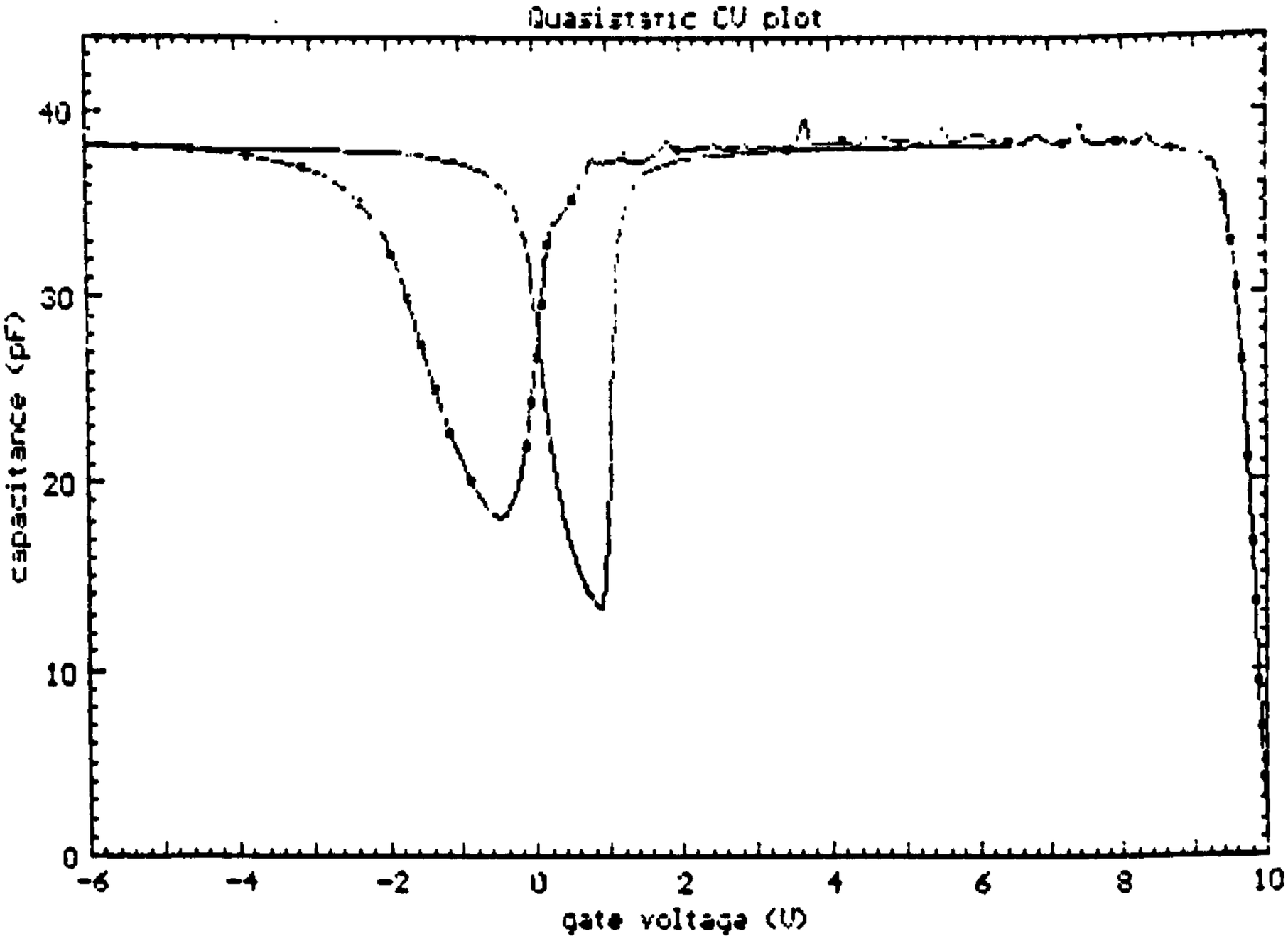
Electron cyclotron resonance enhanced CVD oxide was used for a further bipolar set. This low temperature oxide, and the plasma enhanced CVD oxide grown at Liverpool University for a Zerbst MOS study (discussed in Chapter Four) have resulted from the need to reduce thermal budgets during VLSI processing. A representative quasi static CV plot from a room temperature plasma enhanced oxide (the oxide growth technique is reviewed by Taylor et al, 1993) grown on MBE Si, is shown in Fig. 3.20. By quantitative comparison with the ideal MOS curve, an interface charge density of $1.2x10^{11}$ cm^{-2} is inferred.

In order to planarise mesa trenches i.e. for delta FET gate contacts, polyimide was spun on and annealed at 450 C for 5 minutes. This avoided problems of poor Al

P type Si

dU/dT = -502.1 mV/sec
R = 1e11 ohms

Na = 1.08e16 cm^{-3}
depth = 500 Angstroms
area = 4.91e-4 cm^2
dia = 0.25 mm
Cfb = 29.64 pF
Qo = 8.36e11



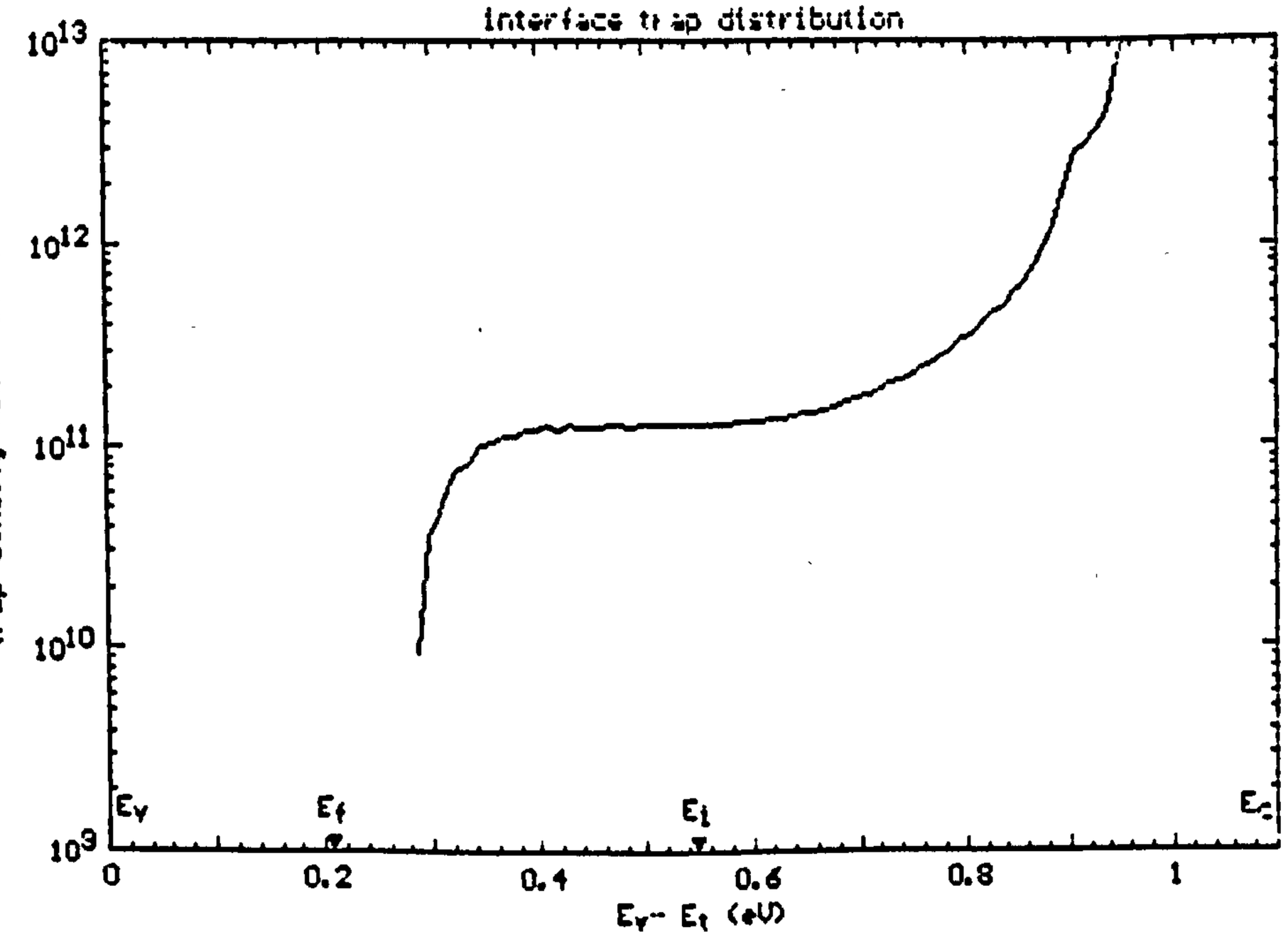
(a)

Only P.A.O. From 10V to -6V.
A34LF6.3 17:37 15-Jun-92

P type Si

Cox = 38.22 pF
D_{it_o} = -6.25e11 cm^{-2} eV^{-1}
$D_{it_{md}}$ = 1.27e11 cm^{-2} eV^{-1}

Na = 1.08e16 cm^{-3}
depth = 500 Angstroms
area = 4.91e-4 cm^2
dia = 0.25 mm
Cfb = 29.64 pF
Qo = 8.36e11



(b)

Only P.A.O. From 10V to -6V.
A34LF6.3 17:33 15-Jun-92

Fig. 3.20 (a) Quasi-static C-V plot obtained from a low temperature plasma enhanced oxide on an Si MBE epilayer, the ideal plot is shown for comparison. (b) The density of interface states inferred from this plot

coverage in the trench and thus track breakage. This was found to be successful, but unfortunately the polyimide produced at the EMF, whilst a good dielectric, does not passivate Si adequately.

3.7 CONCLUSIONS

In-house processing techniques and masks have been developed to enable the rapid production of mesa Hall bars. These techniques and masks have been modified to achieve self-aligned Schottky gated mesa Hall bars at the SUCEM. Ohmic contacts have been made to buried layers by Al deposition and suitable choice of anneal temperatures and times. Processing thermal budgets were less than the growth thermal budgets. Good Schottky contacts have been made to MBE grown Si only after modifications to the growth system designed to reduce metal contamination. Mask sets for processing many analytical and device structures at the EMF have been designed, as have many process sequences shown to preserve the integrity of Si MBE structures. A 1000 C, 15 second furnace anneal has been shown to be an adequate implantation anneal to provide ohmic contacts at 4K. It is suggested that the combination of P doped pyrolytic oxides and some device geometries can produce leakage current problems. Other low temperature oxides have been used with alternative structure geometries and good dielectric performance observed.

CHAPTER FOUR

CHARGE CARRIER LIFETIMES

4.1 INTRODUCTION

This chapter discusses the electrical material quality of MBE grown Si epilayers, at room temperature, and issues relating to the measurement of this quality. During operation of any semiconductor device the charge carrier concentration is locally perturbed from a dynamic equilibrium. The processes occurring to restore equilibrium have a significant impact on the device performance, particularly under transient conditions. The dominating process, certainly in Si and Ge, relates to unintentional impurities and crystalline disorder. Thus, a measurement of the processes restoring charge equilibrium, provides information both regarding expected device performance and regarding material quality. These processes are characterised, experimentally, by the carrier lifetimes - generation, τ_g, and recombination, τ_r.

In this study, the aim of measuring lifetimes is to provide evidence regarding the electrical quality of as-grown MBE Si as a function of growth parameters and to monitor the effect of device processing. Both τ_g and τ_r are important device performance parameters and are particularly interesting because lifetimes are less controllable than other performance parameters, e.g. doping profiles, and can vary by orders of magnitude. The generation lifetime, τ_g, dictates the leakage current of bipolar devices and the transient behaviour of MOS devices, e.g. the refresh time of dynamic RAM. The recombination lifetime, τ_r, governs switching times in bipolar devices.

Any sample process treatment, particularly one necessitating high temperatures, can alter material quality. This is a considerable restriction on the experimental options, which is discussed in more detail after the theoretical background is outlined. This has also been a significant factor in previous measurements of generation lifetime, see Section 4.3. By using a low temperature oxide, values of τ_g have been measured, by an MOS capacitor technique, in V80 grown Si (it is suggested that this particular study is

study is limited by oxide breakdown effects) and in V90S grown Si, both as a function of growth temperature. It is believed that these studies indicate the quality of as-grown Si, in contrast to other reported studies in MBE Si. Further values of τ_g have been obtained, from pn diode measurements, in Si subject to the processing described in Chapter Three. These studies provide values for τ_g in V80 Si, grown at 700 C, and in V90S Si as a function of growth temperature. Recombination lifetimes have been measured in processed diodes, but these values are believed to be erroneous, although a lower limit has been found: this will be discussed briefly. This study was subject to limitations of MBE Si sample supply and of limited access to the EMF.

Recombination and generation processes are outlined, particularly as they relate to material quality. A clear distinction is made between each of these lifetimes and the extrinsic minority carrier lifetimes. A number of features of recombination generation, (RG), processes in epilayers present considerable experimental difficulties, and these are discussed. Previous measurements of lifetimes in MBE Si are few and will be presented in Section 4.3. In order to preserve the fundamental distinction between τ_r and τ_g, the experimental techniques, and results for each, will be dealt with separately; firstly, τ_g, and then τ_r.

4.2 THEORY

The minority carrier lifetimes are a common measure of the processes that restore carrier concentration equilibrium following a perturbation. The concept of lifetime is straightforward: τ_r can be defined in p-type material as

$$\tau_r = \frac{n}{G} \qquad \textbf{4.1}$$

where n is the equilibrium electron density, G the equilibrium generation rate.

The generation lifetime can be similarly defined as

$$\tau_g = \frac{n_i}{G} \qquad \textbf{4.2}$$

Note that τ_g and τ_r are not necessarily equal for fundamental reasons. The time constant measured in a given experiment is often not trivially related to the simple parameters defined above: this will be discussed later. If these lifetimes are to be used as a measure of material quality, then it is important to understand how the measured values may relate to defects. Firstly, the physical processes that can determine the lifetimes will be outlined.

4.2.1 Recombination Generation Processes

There are two classifications of RG: direct and deep level assisted. Direct band-band recombination occurs when an electron drops from the conduction band to valence band, with subsequent loss of energy via radiation and, possibly, phonon emission. Alternatively, an exciton may be formed, particularly at low temperatures, which then decays by radiation or is bound to an impurity and decays with, or without, TO phonon emission. This phenomenon is the basis of photoluminescence studies. Auger recombination refers to the transfer of the lost energy - during band to band recombination - to another charge carrier, which then emits phonons. This is significant in heavily doped material, particularly direct band-gap semiconductors. Alternatively, the energy may be lost to phonons alone in materials with small band-gaps.

Si and Ge have indirect band-gaps; thus, band to band recombination must be phonon assisted in order to provide the necessary conservation of momentum. This, and the relatively large band-gap, reduces the probability of band to band transition via radiative means, such that these processes are not experimentally significant; hence, lifetimes in Si are much longer than in direct gap semiconductors. Indeed, the inability of Si to emit light has been a prime motivation for the development of III-V optical devices, e.g. semiconductor lasers etc. Deep level assisted RG, however, is of paramount importance in Si and Ge.

A deep state in the band-gap may capture an electron and then a hole, i.e. recombination. Such a deep state cannot be described via the effective mass

approximation, because the ionisation energy is of the order of the band-gap. Therefore, these states are defined in terms of coefficients that describe a recombination process, but do not uniquely define the state. These coefficients are the hole (electron) emission and capture rates, e and c, which relate to the capture cross-section, σ, and mean thermal velocity, $\langle v \rangle$, by $c_p = \sigma_p \langle v \rangle p(t)$, i.e. c and e are charge carrier concentration- dependent (and often, therefore, are time-dependent in an experiment).

The details of recombination processes still present considerable theoretical difficulties and will not be discussed in any detail here. Excellent introductions to the theoretical aspects of radiative and deep level recombination are given by Mott, 1978, and Queisser, 1978, respectively.

4.2.2 Recombination

Before obtaining a formalism for recombination and generation at deep levels, it is useful to establish the general expression for recombination. This can be best derived via band to band processes.

At thermal equilibrium, recombination rate, R, and generation, G are by definition equal. We assume that $R \propto np$, hence

$$G = R = Pn_0p_0 = Pn_i^2 \quad , \text{ where P is a rate constant (units } cm^3s^{-1}) \qquad \mathbf{4.3}$$

If equilibrium is disturbed by the injection of excess carriers (Δn Δp) at time t=0, with excess generation rate G_Δ , then

$$G + G_\Delta = P(n_0 + \Delta n)(p_0 + \Delta p) \qquad \mathbf{4.4}$$

Restoration of equilibrium by electron-hole recombination must satisfy;

$$\frac{\partial \Delta n}{\partial t} = \frac{\partial \Delta p}{\partial t} = -G_\Delta \qquad \mathbf{4.5}$$

This is a non-trivial equation. However, it can be tackled by assuming that $\Delta n(t) = \Delta p(t)$. (This can be justified since electrons and holes are generated and recombine in pairs, with the exception of trap events). When $\Delta n = \Delta p$, we can write **4.5** as

$$-\frac{\partial \Delta p}{\partial t} = P(n_0 + p_0 + \Delta p)\Delta p \qquad \mathbf{4.6}$$

i.e. $$\int_0^t \frac{\frac{\partial \Delta p}{\partial t}}{(n_0 + p_0 + \Delta p)\Delta p} dt = -P\int_0^t dt \qquad 4.7$$

The left hand side is an exact differential:

$$\frac{1}{n_0 + p_0}\left[\ln\frac{\Delta p}{n_0 + p_0 + \Delta p}\right]_0^t = -Pt \qquad 4.8$$

Hence, $$\Delta p(t) = \frac{(n_0 + p_0)\Delta p(0)}{(n_0 + p_0 + \Delta p(0))\exp\left(P(n_0 + p_0)t\right) - \Delta p(0)} \qquad 4.9$$

We can now take two limiting cases of **4.9**.

Low Injection

The first case is low injection, i.e. $\Delta p(0) << n_0 + p_0$; so, $\Delta p(t) = \Delta p(0)\exp\left(-P(n_0 + p_0)t\right)$. Defining the excess hole recombination lifetime, τ_p,

as $$\tau_p = \frac{1}{P(n_0 + p_0)} \qquad 4.10$$

allows **4.9** to be written $\Delta p(t) = \Delta p(0)e^{\frac{-t}{\tau_p}}$. Thus, for low injection, recombination precedes by exponential decay.

High Injection

The second case is high injection, i.e. $\Delta n(0) = \Delta p(0) >> n_0 + p_0$, hence

$\Delta p(t) = \frac{\Delta p(0)}{1 + Pt\Delta p(0)}$. Recombination procedes by hyperbolic (or quadratic) decay.

Further, by definition $\frac{\Delta p}{\tau} = \frac{-\partial \Delta p}{\partial t}$, so, from the above, $\tau = \frac{1}{P\Delta p(t)}$. Thus, τ_r is a function of time until the low injection condition is met. Identical arguments, of course, apply for electrons, since we have assumed $\Delta n(t) = \Delta p(t)$.

The most important point to emerge, relevant to an experimental study, is that recombination cannot be assumed to be an exponential decay process nor τ_r a constant. We can now establish the parameter P, for the case of interest.

4.2.3 Deep Level Recombination

A mid gap state (MGS) may either capture an electron and a hole, i.e. it acts as a recombination or generation centre (RGC), or it may capture an electron which may be subsequently emitted to the conduction band i.e. it acts as a trap. For a state to act as a RGC it must be able to localise electrons and holes. The parameters describing an RGC are related to the extrinsic lifetime by Schockley-Read-Hall (1957) statistics. For an RGC, density N_T, of energy E_T, this yields an expression for P;

$$P = \frac{1}{\tau_{p0}(n + n_t) + \tau_{n0}(p + p_t)} \qquad \textbf{4.11}$$

where $n_t = n_i e^{\left(\frac{E_t - E_i'}{kT}\right)}$ and $p_t = n_i e^{\left(\frac{E_i' - E_t}{kT}\right)}$ and $E_i' = E_i - kT\ln g$ to account for degeneracy (E_i is the intrinsic Fermi level).

We define the extrinsic lifetime as $\tau_{p0} = \frac{1}{\sigma_n v_{th} N_t}$ **4.12**

where v_{th} is the carrier thermal velocity ($= \sqrt{3kT/m}$) and σ is the capture cross section, which can vary from ~10^{-16} cm^{-2} for surface states to ~10^{-10} cm^{-2} for Be.

Hence, from 4.11 and the definition of τ_p (τ_n) in **4.10**

$$\frac{1}{\tau_p} = \frac{1}{\tau_n} = \frac{n_0 + p_0 + \Delta p(t)}{\tau_{p0}(n_0 + n_t + \Delta n(t)) + \tau_{n0}(p_0 + p_t + \Delta p(t))} \qquad \textbf{4.13}$$

There are three consequences of this expression that have important implications for the relationship between τ_r, MGS and τ_{n0}, τ_{p0}:

(i) Deep levels with E_T near midgap will yield the maximum rate of recombination (assuming equal σ and N_t). This effect is important for generation where the effect of states away from mid-gap drops off exponentially (see following Section).

(ii) If there is an excess of holes (i.e. p-type material, $p_0 > n_0$), then the recombination rate is limited by minority carrier (electron) capture rate, since $P \approx \frac{1}{\tau_{n0}(p + p_t)}$.

Therefore, $\tau_r = \tau_n$, in the absence of trap effects.

(iii) If, and only if, $E_t = E_i$ then $\tau_n = \tau_{n0}$, since $p_t = 0$ and $\tau_n = \frac{n}{Pn_i^2}$

Proviso

The forgoing is appropriate only when the deep state density $N_t < \Delta n, \Delta p$. In the steady state i.e. $G = \frac{\Delta n}{\tau_n} = \frac{\Delta p}{\tau_p}$, then $\frac{\tau_p}{\tau_n} = \frac{\Delta p}{\Delta n}$. In the case where N_t is large, then $\tau_p \neq \tau_n$, since charge neutrality requires that the number of occupied deep states, n_t, is introduced (e.g. $\Delta p = \Delta n + \Delta n_t$, for deep acceptors), therefore, $\tau_p \neq \tau_n$.

4.2.4 Generation

It is appropriate to establish a formalism for τ_g. Net generation occurs usually in a depletion region i.e. where n, p = 0. Since $G = Pn_i^2$ and $\tau_g = n_i/G$ then

$$\frac{1}{\tau_g} = Pn_i \qquad 4.14$$

Taking the SRH expression for P (equation **4.11**)and setting n = p = 0 yields

$$\tau_g = \tau_{p0} e^{\frac{E_t - E_i'}{kT}} + \tau_{n0} e^{\frac{E_i' - E_t}{kT}} \qquad 4.15$$

When $E_i = E_t$, then τ_g will represent the longer of the extrinsic lifetimes. For $E_t \neq E_i$, (provided that $|E_t - E_i| > kT$), τ_g will be longer than either τ_{p0} or τ_{n0}, quite possibly by orders of magnitude. The link to recombination lifetime is, therefore, even more tenuous. This is illustrated in Fig 4.1. Note that τ_g samples a narrow region of the band gap only, and that τ_r is more sensitive to mid gap states.

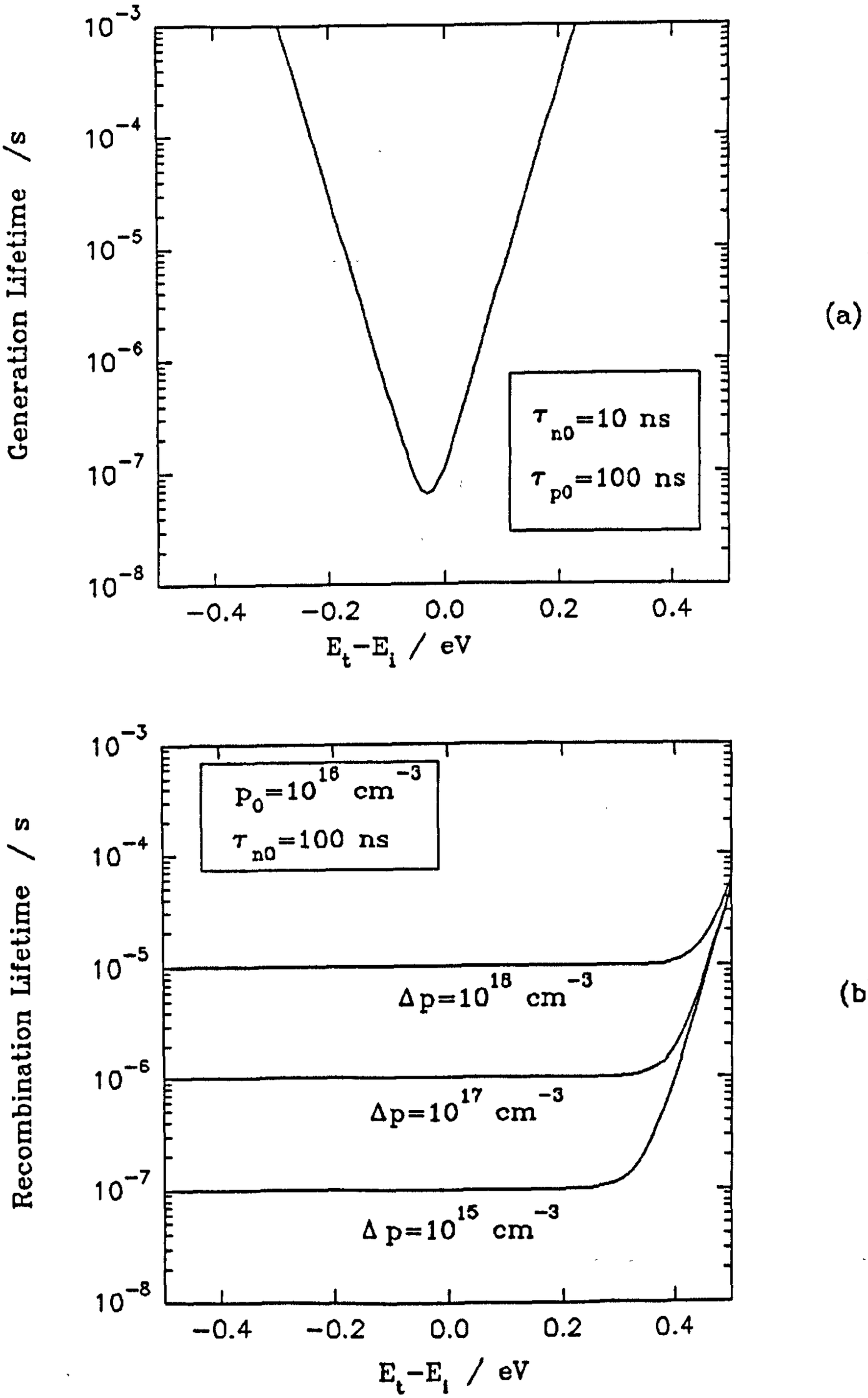


Fig. 4.1 Illustration of the sensitivity of generation (a) and recombination (b) lifetimes to midgap state energy level with respect to midgap. Both low and high injection conditions are shown in (b)

The parameter τ_g is of considerable practical importance for MOS devices and the reverse current of pn diodes. It is also occasionally quoted as the minority carrier lifetime in material quality studies, although, of course, this is rarely the case. The particular significance of τ_g is that it is relatively easy to obtain a valid measurement in epitaxial layers (provided that a high quality dielectric is available), whereas measured values of τ_r are often erroneous from thin layers. This will be discussed in section 4.6.

4.2.5 Surface Effects

The most important consideration for measurements of τ_r or τ_g, excepting the lifetime itself, is the effect of any surface (or interface, e.g. a pn junction). The recombination rate at a surface is likely to differ from that in the bulk, due to surface states etc.

The surface is characterised by a constant of proportionality that relates the excess carrier concentration to the absorption rate. For hole concentration, p, with mean thermal velocity v, the number of holes per second impinging on unit area of a surface is vp/4. If r is the probability of reflection at the surface and S_e the number emitted per second, then, in equilibrium,

$$\frac{vp_0}{4} = \frac{vp_0}{4} r + S_e \qquad \text{, i.e.} \quad S_e = \frac{vp_0}{4}(1-r) \qquad \textbf{4.16}$$

In non-equilibrium, $p = p_0 + \Delta p$, therefore, the rate of absorption of excess carriers per second, S_a, is

$$S_a = \frac{v}{4}(1-r)\Delta p \qquad \textbf{4.17}$$

Thus, the constant of proportionality has unit ms^{-1} and is termed the surface recombination velocity, s.

A system with bulk generation rate of excess carriers, G_e, is governed by

$$D\frac{d^2\Delta p}{dx^2} = \frac{\Delta p}{\tau_p} + G_e \qquad \textbf{4.18}$$

If the surface is at x = 0, then two boundary conditions are:

$\Delta p \rightarrow G_e\tau_p$ as $x \rightarrow \infty$, and $D\frac{\partial \Delta p}{\partial x} = s\Delta p$ at x=0.

This is readily solved to yield

$$\Delta p = G\tau_p\left(1 - \frac{s\tau_p e^{-x/L}}{L + s\tau_p}\right) \qquad \text{, where L is the diffusion length, } \sqrt{D\tau_p}\,. \qquad \textbf{4.19}$$

This has two consequences. Firstly, that the effect of the surface has exponential decay with characteristic length governed by the ratio x/L. Secondly, at the surface, if $s >> D/L$, then $\Delta p(0) \rightarrow 0$ (i.e. total absorption), but for $s < D/L$, then $\Delta p(0) \rightarrow G_e\tau_p$ (the bulk value). A normalised surface recombination velocity, S, is defined for convenience, $S = \frac{sL}{D}$. So, S=0 implies perfect reflection at a surface, $S >> 1$ implies a perfect sink.

It is clear that experimental techniques that are either unaffected by s, or that measure s, are highly desirable. This latter requirement is complicated by the fact that s can vary during an experiment. This is discussed by Rees, 1985, De Visschere, 1986, and latterly by Correig et al 1990. The variation of s was observed experimentally in MOS capacitor measurements by Schroder and Nathansen, 1970. The effect of this during experiment was first noted as a correlation between high values s and apparently low τ_g. The variation of s in transient measurements is hardly surprising from physical considerations: as an MOS capacitor is pulsed from accumulation to inversion, most surface R-G centres will be swept through the Fermi level, thereby emptying and filling at non-equilibrium rates. Since surface recombination can be described as SRH processes occurring via surface states, s must

vary considerably. This effect has been mitigated where possible in this study (see. Section 4.5.1).

4.2.6 Trap Effects

As further illustration of the potential difficulties in interpreting lifetime measurements, a brief outline of the influence of traps is useful. If a deep level is close in energy to a band edge then it may capture and emit to that band i.e. without recombination. Such a level is described as a trap. We are not concerned with trap concentrations per se, however, trap effects may result in an erroneous value for τ_r experimentally.

The general expression for decay processes in the presence of a significant trap concentration is unwieldy. Two simple cases illustrate the danger of ignoring trap effects. Firstly, when the recombination rate is much higher than the trap emission (τ_e) and capture (τ_c) rates. Here, the free carrier concentration, n, can be written

$$n = n^0\left(\frac{\tau_n}{\tau_e}e^{\frac{-t}{\tau_e}} + e^{\frac{-t}{\tau_n}}\right) \qquad 4.20$$

The effect of this is illustrated in Fig. 4.2. Clearly, problems can arise if $\tau_n < \tau_e$, τ_c; so the short initial recovery rate (governed by τ_n) may not be observed experimentally or may be attributed to a surface effect.

If the trap density is high and $\frac{1}{\tau_c} >> \frac{1}{\tau_n}, \frac{1}{\tau_e}$ then $n = n^0 e^{\frac{-t}{\tau_{sum}}}$ 4.21

where $\tau_{sum} = \tau_n\left(\frac{\tau_e}{\tau_c}\right)$. Again, an exponential decay can yield a misleading value for τ_n. If the recombination and trapping rates are of the same order, then a non-

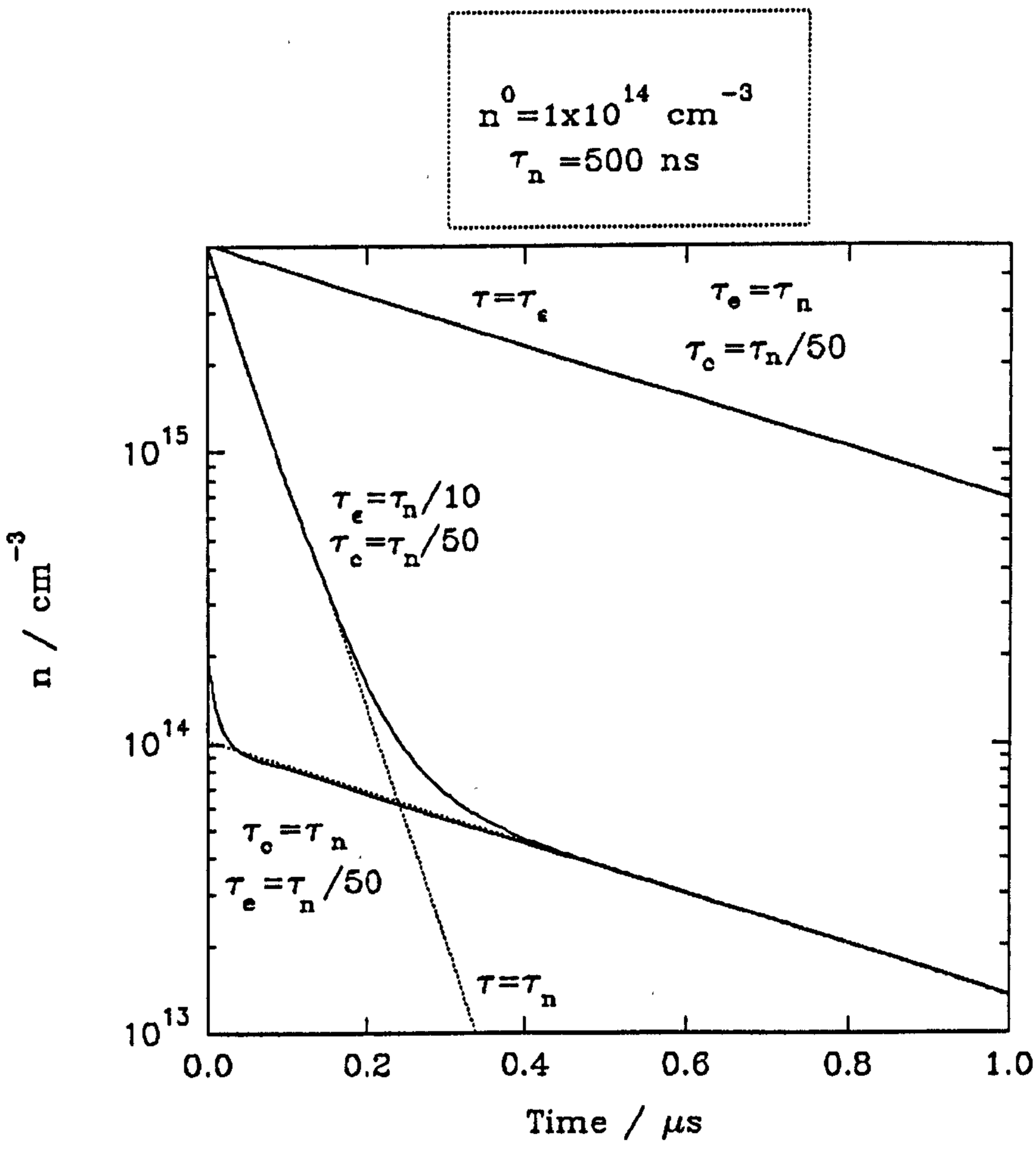


Fig. 4.2 Representative excess carrier decays with time in the presence of significant trap concentrations

exponential decay results, which is easier to notice experimentally, but difficult to interpret.

It can be argued that whether a decay rate is determined by trap effects or deep level recombination is unimportant in device operation: it is only necessary to know the characteristic decay time constant. However, for recovery governed by **4.20**, the possible neglect of the initial rapid recovery can lead to predictive models that are at variance with real behaviour over the full range of operating conditions. For large trap concentrations, yielding the decay of **4.21**, a relatively slow decay, i.e. $\tau_e < \tau_c$, may obscure deep level contamination which could otherwise be reduced.

4.2.7 Defects, Deep Levels And Doping

The relationship between defects and R-G centres have been extensively studied. Despite the vast literature on the subject, however, there are many examples of conflicting evidence about the role of point and extended defects (some of which were due, in early studies, to residual metal contamination). Some of the most effective experimental approaches are the SEM based techniques such as electron beam induced current, EBIC, reviewed by Ravi 1981, and capacitance transient techniques, e.g. DLTS. A review of this field is beyond the scope of this study. The difficulty of attributing R-G centres to a particular source was discussed in Chapter 2. In this study, no attempt is made to firmly associate a lifetime with a particular defect or contaminant.

Experimentally, the highest values of τ_r are found to be approximately constant for $N_{d,a} < 10^{16}$ cm^{-3} and to decrease monatonically with doping concentration above this value, largely due to Auger effects. Fossum and Lee 1982 have modelled the highest reported experimental values of τ_r as a function of doping density by assuming a single acceptor-like defect. Although this modelled the lifetime results, there is little

experimental evidence for such a dominant defect, for instance, two donor-like defects have been reported in high quality substrates by Yau and Sah, 1974.

4.3 PREVIOUS WORK

To the authors knowledge, no systematic study of lifetimes in as-grown MBE Si as a function of growth parameters has been reported. There are, however, a number of reported values of τ_r and τ_g, although the distinction is occasionally not made. A brief summary of these values is given here, as are all available growth and measurement details. It is not sensible to be dogmatic about what constitutes good or bad material quality. However, to put these values in context, measured τ_g in polysilicon are of the order of nanoseconds, in Si on insulator around 10 ns, in substrates generally $\geq$10 μs, in gettered devices some 10's of milliseconds, while the maximum value of τ_r in Si is approximately 1 ms. High temperature processed devices are expected to yield τ_g of 100 μs.

Generation lifetimes have been measured using MOS capacitors by Ota 1983, Peaker et al 1986, and Liu 1988. Ota used high temperature oxidation (1050C 60 mins) to form the dielectric and observed τ_g=15 μs in both MBE Si (T_s=850C) and CVD grown Si. This uniformity might suggest τ_g defined by the oxidation process (note the use of an approximate analysis that neglects surface effects - see section 4.4.2). Peaker et al also used high temperature oxidation and uniformly measured τ_g=30 ns in MBE Si supplied by various growth centres. DLTS results taken prior to oxidation had predicted τ_g from 18 ns to 1 μs. Again, high temperature oxidation would appear to be influential. Liu measured a τ_g =20 ns in Si, grown by MBE, which was later found to contain Ta at 3 x 10^{16} cm^{-3}, Cu at 3 x 10^{15} cm^{-3} and W at 3.4 x 10^{14} cm^{-3}. No further details are available.

Generation lifetimes have been inferred from pn diode leakage currents by Ota 1983 and Becker and Bean 1977. Ota used high temperature processing to form the diode, in an n- epilayer, comprising both B diffusion and steam oxidation, and found τ_g=70 μs. Becker and Bean measured τ_g=100 μs from unpassivated pn diodes with the junction formed at the epilayer(p)/substrate(n^+) interface. (No further details of diode formation are given.) The interface region, therefore, was within the depletion region and will contribute to generation. Both of these measurements assumed the reverse current to be given by qn_iW/τ_g.

Recombination lifetime measurements have been reported by Grivitskas et al, 1988, and Higashi et al, 1990. The former used a transient grating technique to separated the SRH lifetimes from Auger effects and found τ_r=0.5 ns and 10 ns in heavily doped (10^{20} cm^{-3} and 10^{17} cm^{-3} respectively) Si grown in a V80. The latter reports continual monitoring of τ_r in as-grown Si, by placing the sample in the magnetic field of an inductor forming part of a resonant circuit. An excess of minority carriers is introduced by optical stimulation from a strobe lamp, the phase shift resulting in the circuit is then measured as a function of time. They report $\tau_r \leq 100$ ns, in Si (T_s = 750 C) grown after an Ar sputter substrate clean, and τ_r of 50 to 200 μs, in Si (T_s = 750 C) grown after an ex-situ HF substrate clean. Rotation was reported to "kill the lifetime", even without the Ar sputter clean: the longer lifetimes were found in Si grown without rotation. Note that a Ta heater was used in these growths. Sample processing, comprising implantation and activation anneals, is reported to either reduce τ_r to 100 ns, or to have no effect, depending on the processor. Surface effects are neglected: the author suggests that an HF dip is adequate to eliminate surface effects. It is not clear why the problems of minority carrier diffusion to the interface and the substrate are not factors in these results (the 1/e carrier decay length for τ_r=250 μs is of the order 100's of microns, see the Section 4.6).

It is clear that measured lifetimes in MBE Si vary over at least three orders of magnitude. It is stressed, that the thermal history of a sample prior to measurement may alter the lifetimes, by orders of magnitude. The roles of gettering - which may be intrinsic - and post growth contamination, in the values measured in MBE Si, may be important. As previously mentioned, effective gettering can increase τ_g by at least three orders of magnitude. However, no effective gettering process has been reported, that is compatible with low dimensional structures. Lifetimes in as-grown Si, therefore, where reported, are generally similar to those found in highly disordered Si, with the exception of material grown by Bell Labs (Higashi, Becker and Bean), which, it appears, can be of high quality, although it is not clear why this is the exception.

There is clearly a need for systematic measurements of lifetimes in MBE grown Si, since lifetimes of 10's ns are not adequate for commercial exploitation and indicate serious problems with some aspect of MBE growth that must be rectified, if possible.

4.4 GENERATION LIFETIME - EXPERIMENTAL TECIINIQUES

4.4.1 Introduction

It is apparent that, from the previous outline of recombination generation theory, the accurate experimental determination of lifetime requires great care. This is particularly true for measurements of τ_r in epitaxial layers (see Section 4.6). The generation lifetime, τ_g, is much more straightforward to measure since generation will occur in a depletion region, which may be readily confined to the epilayer. Thus, techniques developed for bulk material can be directly applicable to MBE Si.

4.4.2 Capacitance Transients

The extraction of τ_g from MOS capacitor measurements was elucidated by Zerbst, 1966. The theory of the MOS capacitor was outlined in Section 3.5.2. If the capacitor is biased to accumulation, then pulsed into strong inversion, the depletion width, W, that results will be initially greater than the equilibrium depletion width, W_f, in order to compensate the gate charge. The relaxation of the depletion width to equilibrium can be observed as a function of time - the capacitance transient (C-t). The restoration of equilibrium occurs primarily by the generation of electron hole pairs in the depletion width. Minority carriers are swept to the Si/SiO_2 surface (by the field in the depletion region) and enter the inversion layer. Majority carriers are repelled out of the depletion region and either contribute to the current flow in the external circuit or neutralise ionised impurity atoms to reduce W. Fig. 4.3 a & b.

At any point, the charge on the gate Q_g must be compensated by the charge in the Si due to inversion, Q_{inv}, and depletion, so

$$Q_g = Q_{inv} + qN_aW \qquad \textbf{4.22}$$

and by Gauss's law $$Q_g = C_0\left(V_g - \phi_s\right) \qquad \textbf{4.23}$$

Equating **4.22** and **4.23**, (with V_g constant for t≥0) we can differentiate to find an expression for the transient.

$$-C_0\frac{d\phi_s}{dt} = \frac{dQ_{inv}}{dt} + qN_a\frac{dW}{dt} \qquad \textbf{4.24}$$

Now $\phi_s = \frac{qN_aW^2}{2\varepsilon_0\varepsilon_s}$ and, including surface effects , $\frac{dQ_{inv}}{dt} = \frac{qn_i}{2\tau_g}\left(W - W_f\right) + qn_is$

So **4.26** becomes

$$\frac{dW}{dt}\left(qN_a + \frac{C_0qN_aW}{\varepsilon_0\varepsilon_s}\right) = \frac{qn_i}{2\tau_g}\left(W - W_f\right) - qn_is \qquad \textbf{4.25}$$

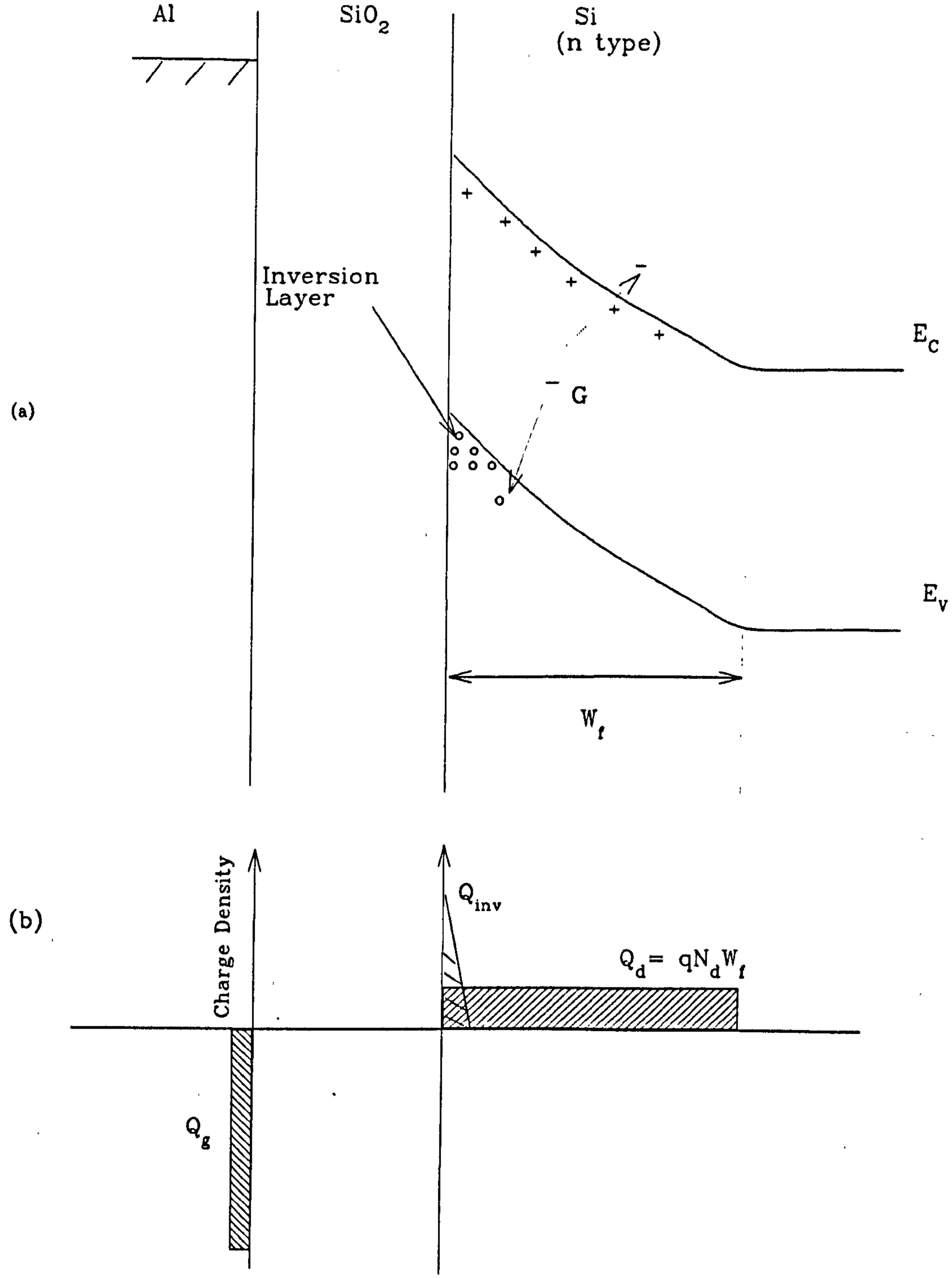


Fig. 4.3 The generation process and charge distribution during a capacitance transient.

If $W = \varepsilon_0 \varepsilon_s \left(\frac{1}{C} - \frac{1}{C_0} \right)$, then this yields $\frac{C_0}{C^3} \frac{dC}{dt} = \frac{n_i}{N_a \tau_g} \left(\frac{1}{C} - \frac{1}{C_f} \right) + \frac{n_i s}{\varepsilon_0 \varepsilon_s N_a}$ **4.26**

which may be rewritten as

$$-\frac{d\left(\frac{C_0}{C}\right)^2}{dt} = \frac{2n_i}{\tau_g N_a} \frac{C_0}{C_f} \left(\frac{C_f}{C} - 1 \right) + \frac{2\varepsilon_{ox} n_i s}{\varepsilon_s t_{ox} N_a} \qquad \textbf{4.27}$$

This is the Zerbst equation. A plot of $\frac{d\left(\frac{C_0}{C}\right)^2}{dt}$ versus $\left(\frac{C_f}{C} - 1 \right)$ will yield gradient proportional to $1/\tau_g$, intercept proportional to s, with all other parameters known. There are two important features of this equation. The first relates to the length of the transient, the second to the separation of bulk and surface effects. Neglecting surface effects, and solving **4.29**, (setting $C=C_f$), yields length of transient $t_t = \frac{N_a}{n_i} \frac{C_0}{C_f} \tau_g$. For $N_a = 10^{16}$ cm^{-3}, this gives a measurement time of ~$10^6 \tau_g$. So, the range of τ_g that might be encountered - 1 ns to 10 ms - gives measurement times greater than milliseconds; thus, the C-t can be readily measured with relatively simple equipment. The importance of the separation of surface and bulk effects has been made clear in the theory section.

Two practical problems with the Zerbst analysis have been proposed. The first is the need to differentiate the transient; this, however, is no longer prohibitory as a result of modern computer power. The second problem is the length of transient in high quality Si, which can be up to an hour. A number of mathematical approximations of the Zerbst analysis have been developed in an attempt to overcome this difficulty. (See for example Heiman, 1967 or Huang, 1970). An excellent critical review by Kang and Schroder illustrates that such simplifications can yield measured τ_g that are in error by up to an order of magnitude - largely due to the neglect of surface

effects - even for a high quality Si/SiO_2 interface with $s = 0.3\ cms^{-1}$. This review also assesses many of the optional MOS lifetime measurement techniques e.g. measuring current and capacitance simultaneously, using a linear ramp voltage etc. A full discussion of these would be lengthy and could add little to the aforementioned review, which is recommended.

The only difficulty presented by the Zerbst technique for MBE structures is the need for an oxide. As discussed, the growth of a high temperature oxide may well alter τ_g either by introducing contamination or by initiating intrinsic gettering. High temperature oxides are not a practical option for MBE grown structures due to the effect of high temperatures on abrupt dopant profiles or strained alloys.

In this study, for the first time, low temperature oxides were used: specifically, the plasma enhanced oxide grown at Liverpool University. Subsequently, 0.25 mm diameter Al contacts were formed by evaporation through an alumina mask, which was not in contact with the sample. These contacts were given a ten minute 400 C post metallisation anneal in forming gas. Note that this represents the highest post growth thermal budget for the samples. Approximately 100 capacitors were formed on each substrate. These were probed using a Wentworth probe station.

High frequency (1 MHz) and quasi-static C-V plots were obtained using equipment and software supplied by Liverpool University. Pulsed MOS transients were recorded using a Boonton 72B capacitance metre, with a response time of 50μs, and a Tektronix 2430 oscilloscope, Fig. 4.4. The data was down loaded to a PC, converted to ASCII format and then manipulated using Jandell software, Sigma Plot. This manipulation comprises capacitance normalisation to C_0, plotting $(C_0/C)^2$ versus t, fitting the data to a polynomial, manual differentiation of this polynomial, then Zerbst plot generation. This process is illustrated by Fig. 4.5. Note that the use of a polynomial for fitting does not force an exponential form on the data. Also, that no fit

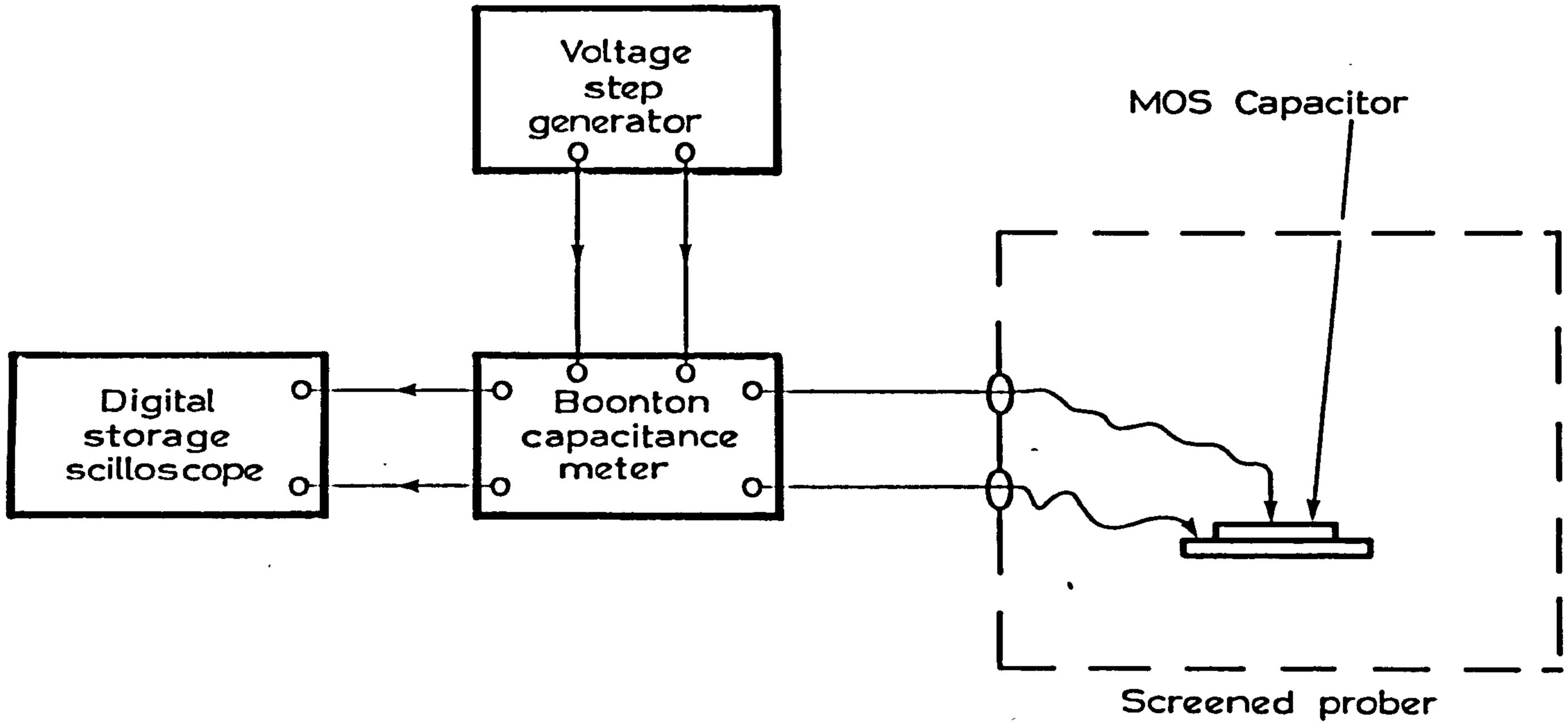
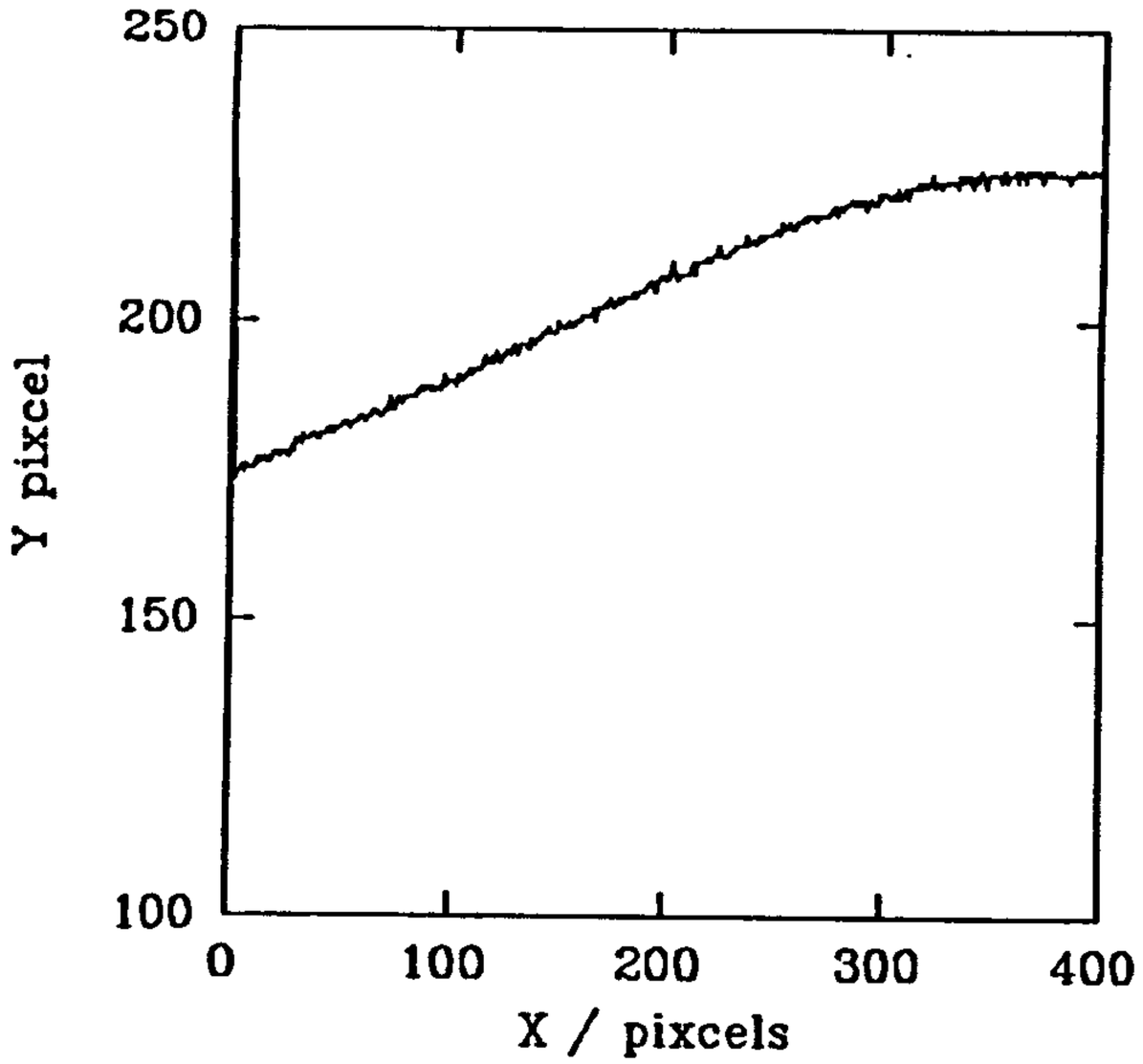
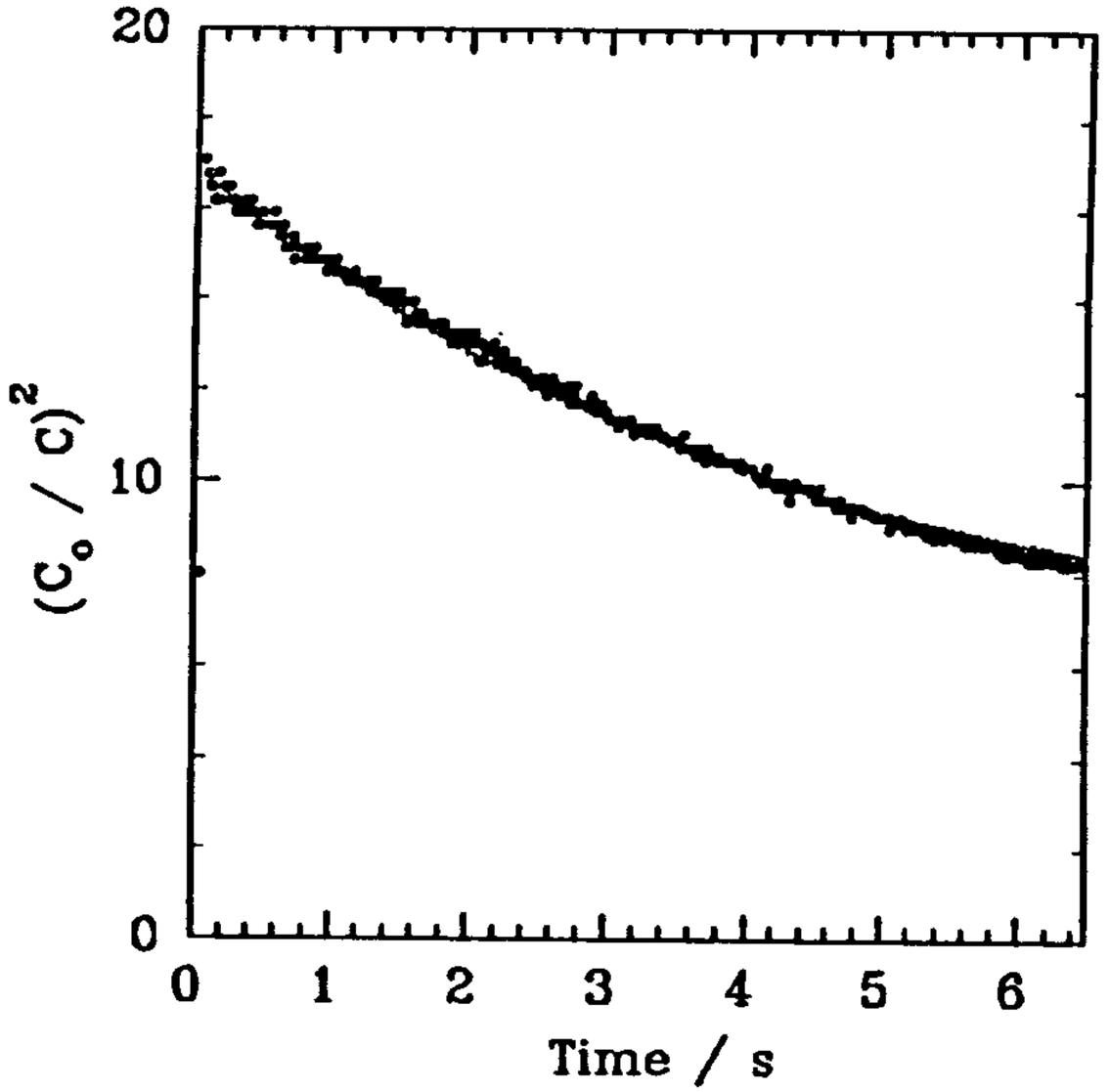
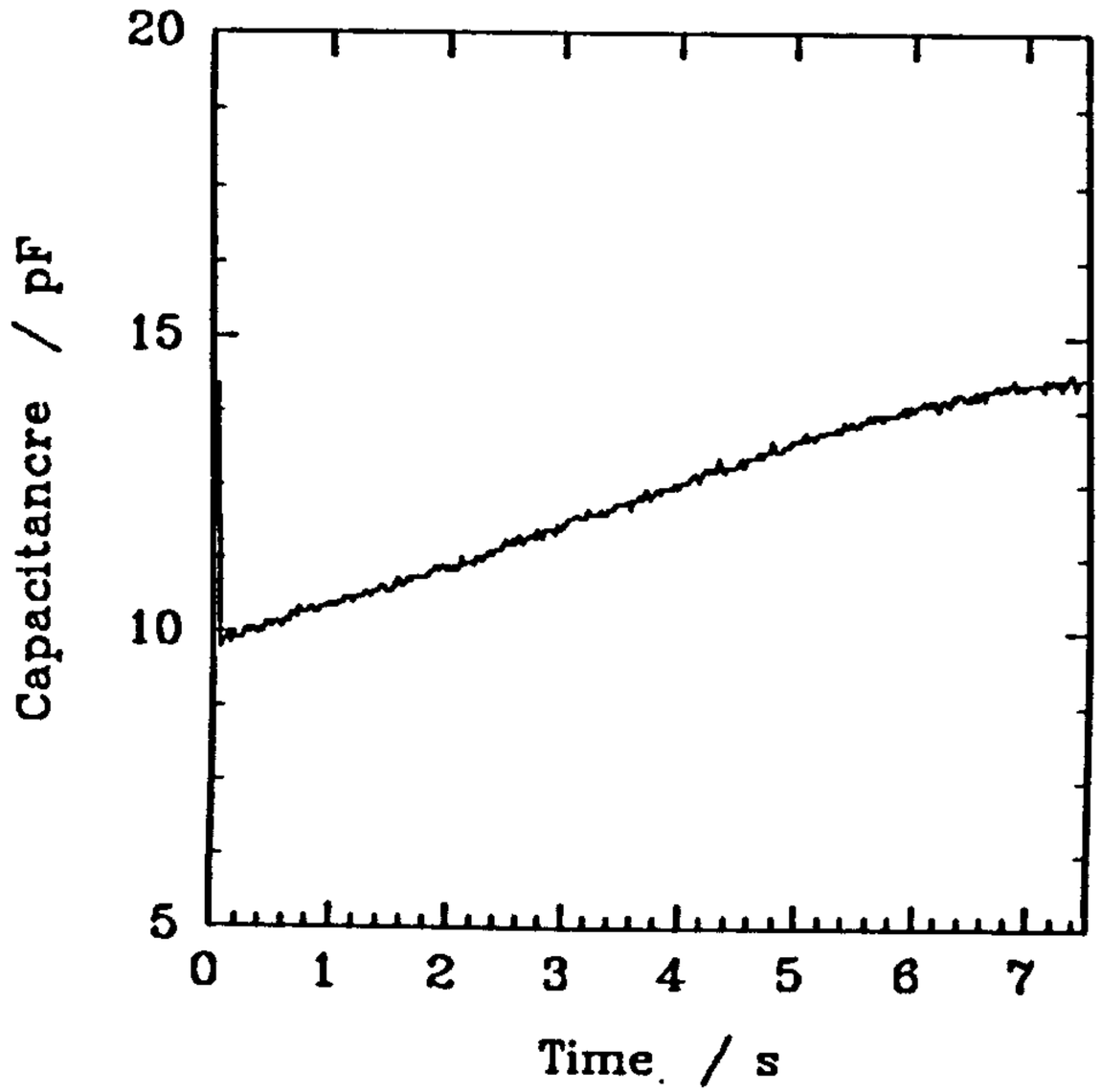


Fig. 4.4 Apparatus used to measure the capacitance transient.







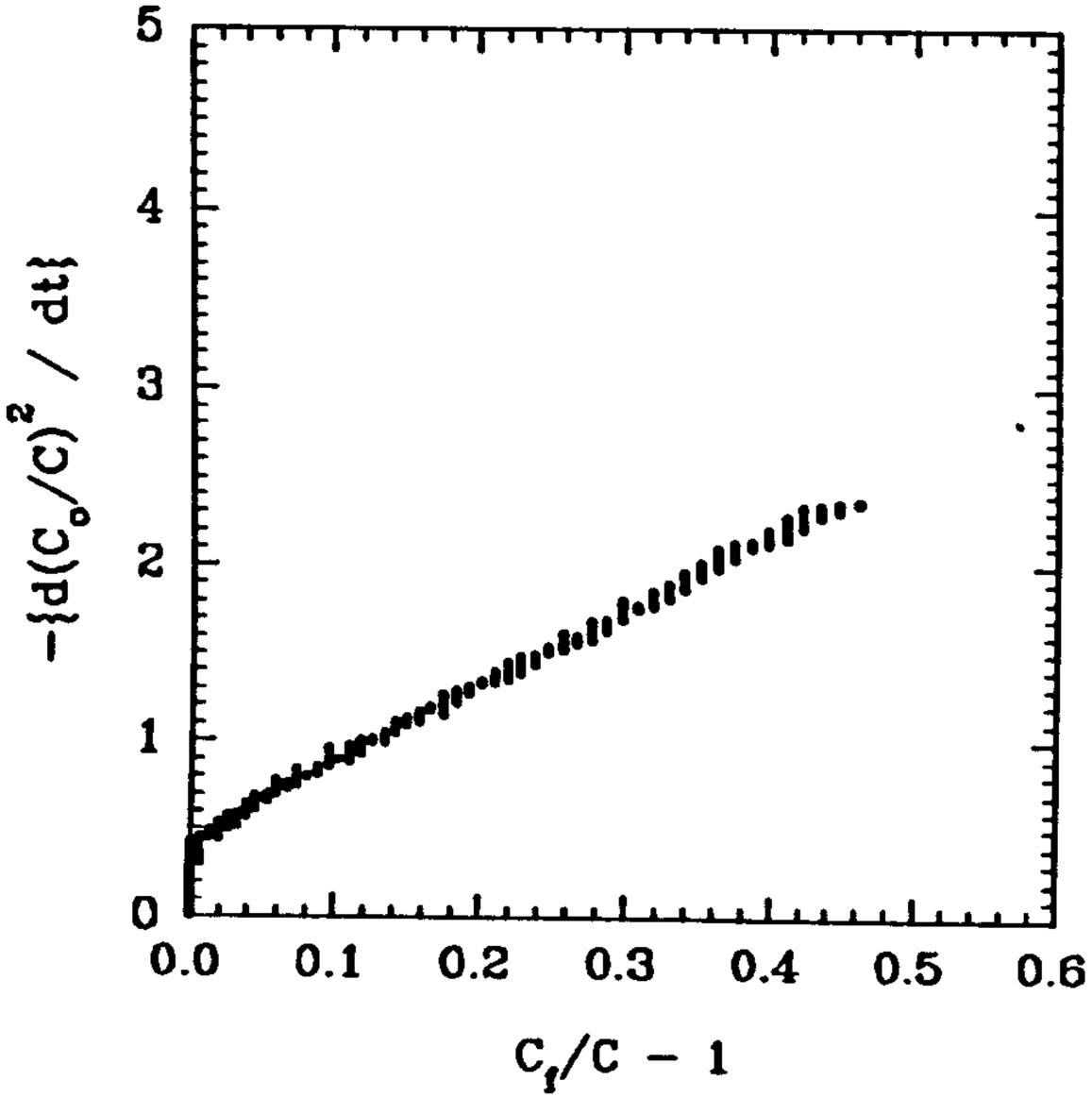


Fig. 4.5 The four analysis stages required to generate a Zerbst plot from the oscilloscope waveform.

was accepted with correlation coefficient less than 0.99. Any linear region in the Zerbst plot, therefore, can be considered a true reflection of the data.

4.4.3 pn Diode Measurements

The Zerbst technique allied to a low temperature oxide allows measurements of τ_g with the minimum post-growth processing and thermal budget. However, the need for a high quality oxide and large area capacitors renders it unsuitable for monitoring VLSI processed samples. To overcome this deficiency the author chose the pn diode as a suitable structure for the measurement of τ_g (and, more importantly, τ_r) in processed structures. This section discusses ways of extracting τ_g from diode measurements.

As previously stated (Section 3.6.5) the reverse current of a pn diode may be written

$J_r = J_{diffusion} + J_{generation}$, where $J_{generation} \propto W \propto V^{1/2}$

(Note that multiplication current effects are avoided, in this study, by analysis at $Vr < V(breakdown)/3$). Thus, from IV data, a plot of I_r versus $V^{1/2}$ should yield straight line with gradient qn_iA/τ_g, so τ_g may be found. However, a surface term, J_{surf}, must be added to this (as the extreme case shown in Chapter 3 demonstrated), $I_{surf} = qn_isA_s$, where A_s is the surface area. For square diodes, side L, therefore, the three components of I_r are $I_{diffusion}$ ($\propto L^2$) , $I_{generation}$ ($\propto L^2$, W), and I_{surf} ($\propto L$). So, measurement of an array of diodes of different areas will enable the presence of significant surface currents to be confirmed, i.e. a plot of $I_r(V)$ against L^2 will yield a straight line in the absence of surface currents.

In light of the previous results, indicating significant surface effects, mask set Eμ 931 includes diodes of side in the range 120 to 300 μm, for this purpose. This, however, does not readily yield τ_g in the presence of high, and possibly varying, surface currents, i.e. non-linear I(V) versus L^2. Further, for high quality devices (τ_g

> 100 μs) and, therefore, low J_{gen}, the diffusion current cannot be neglected, even though it will be a constant. As Schroder, 1984, has demonstrated, order of magnitude underestimates of τ_g can result from a straightforward analysis of the reverse current measurements in high quality material. A solution to this latter difficulty (and the former, for relatively small J_{surf}) is to create a plot analogous to the Zerbst, i.e. of I_r versus W (in the Zerbst plot $C_f/C-1 \equiv W-W_f$). This necessitates a CV measurement and slightly more analysis than required from the usual pn diode measurement, but has the distinct advantage that the correct value of τ_g may be obtained over a wide range of material quality.

In order to find W (V), a high frequency capacitance measurement must be performed which will yield a linear $1/C^2$ versus V_{rev} plot, provided that C is dominated by the depletion capacitance. For a one sided abrupt junction

$$C_{dep} = \frac{-\varepsilon_s A}{\sqrt{2} L_D} \sqrt{\frac{q}{kT}(V_{bi} - V) - 2} \quad \text{and} \qquad \textbf{4.28}$$

$$W = L_D \sqrt{2\left(\frac{q}{kT}(V_{bi} - V) - 2\right)} \quad \text{so,} \qquad \textbf{4.29}$$

$$W(V) = \frac{\varepsilon_s A}{C(V)}, \text{ which is, of course, the capacitance of a parallel plate capacitor.} \qquad \textbf{4.30}$$

Thus, we need only plot I_r versus $\varepsilon_s A/C$ to obtain a quasi Zerbst plot for the pn diode.

IV measurements were performed using two measurements systems in this study: the first constructed by the author and latterly, an HP Parameter Analyser. The former system was controlled by a BBC Master computer with software written by the author. Difficulties due to ammeter input impedance effects were encountered with the first arrangement - a Keithley 485 picoammeter in series with the test device - which resulted in significant loading of the diode. The current was therefore measured using a 100 Ω standard resistor in series with the device, the voltage being measured with a Keithley 181 nanovoltmeter. Devices were probed using a Wentworth probe station.

The aim of these lifetime studies was to provide room temperature values. However, some low temperature measurements were taken on packaged diodes as discussed in Section 3.6.5. These were performed in a liquid N_2 bath cryostat constructed by the author. The temperature was monitored by a thermo couple and the heater unit, also constructed by the author, based upon a Eurotherm 820 temperature controller. The cryostat was rotary pumped to 10^{-2} torr, then continually purged with gaseous He. An ingenious He non-return valve - a beaker of water - was adequate for most of these measurements. However, a cryo-pumping effect, combined with unstable He flow, dramatically led to the abandonment of this cryostat, which was replaced by a closed cycle system described in section 5.3.2.

For completeness, the gated diode is mentioned as an option for τ_g measurements. This structure and lifetime measurements that may be derived from it are fully discussed by Grove and Fitzgerald, 1966. Whilst this is an attractive technique, it requires both diode formation and a high quality oxide. This was felt to entail an excessive process burden in the early stages of this study. It may, however, prove to be a useful technique as low temperature oxides become more widely available.

4.5 GENERATION LIFETIME - RESULTS AND DISCUSSION

4.5.1 Zerbst Analysis

The author has designed and measured two sample sets using MOS capacitors. The first set was grown in the V80, the second in the V90S (after modifications had been made to the Si source designed to reduce metal contamination). This first (V80) sample set provided some useful information, but was not wholly satisfactory due to

oxide breakdown problems: these limited the statistical validity of the study. The later V90S results will be discussed in greater detail.

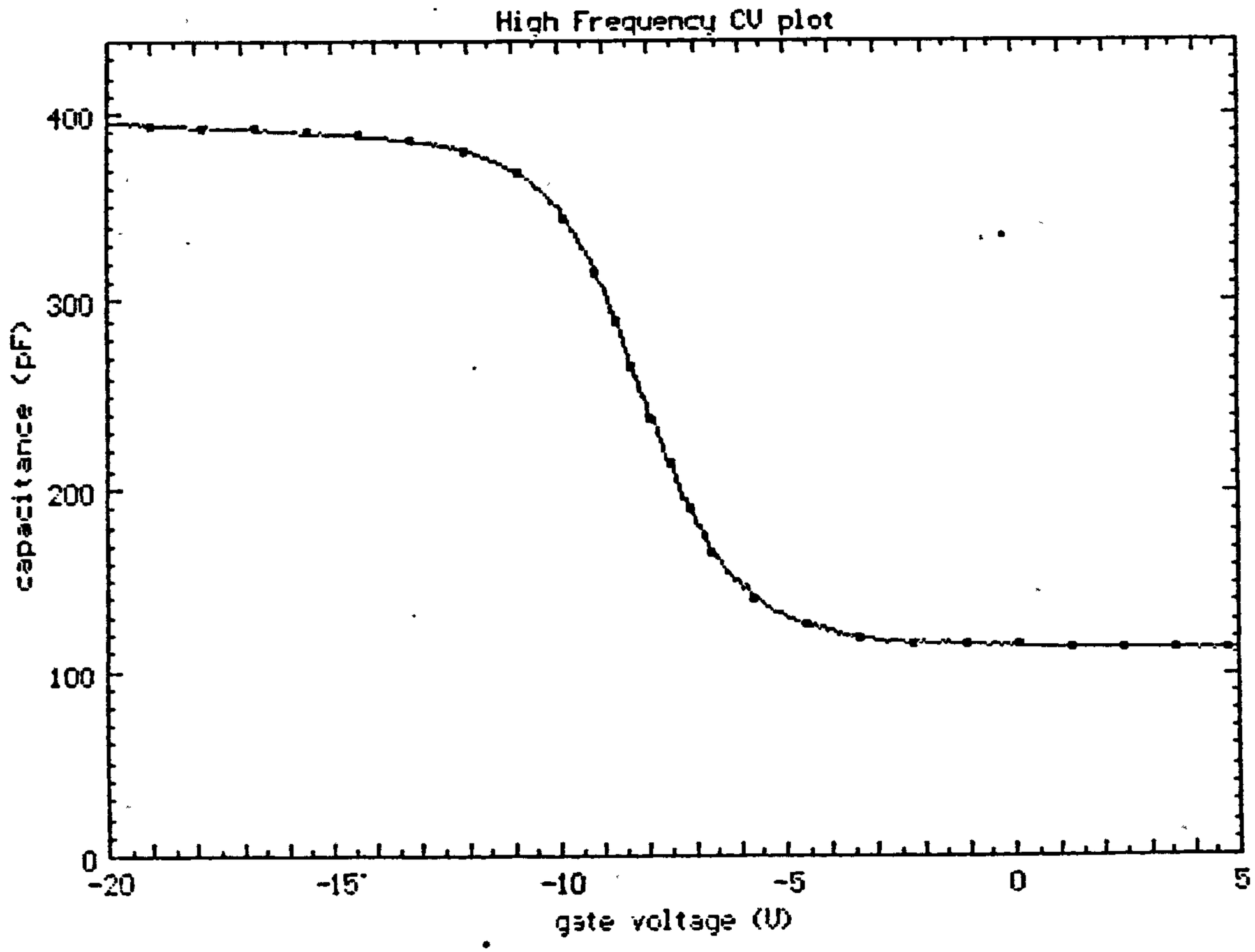
4.5.1.1 V80 Grown Si

The first sample set consisted of 6 epilayers grown at T_s of 500, 600, 650, 700, 750 and 800 C. Growth order was randomised with respect to T_s and a p^- substrate, of nominally same doping concentration as the epilayer, was included in the sample set as a control. The epilayers were B doped, nominally 10^{16} cm^{-3}, to overcome the n type background doping and were grown on 3" Czochralski p^+ substrates. The equilibrium depletion width of an MOS capacitor biased into an inversion with this doping level, is 0.3um. The epilayer thickness was chosen to be 1.5 um. This it was hoped, would potentially allow lifetime profiling of the epilayer up to the substrate/epilayer interface by increasing the deep depletion pulse. Note that it is possible to find τ_g in the case of non-uniform doping concentration (Miyake and Harada, 1981), e.g. in the B interface spike. The extent of deep depletion will be limited by the oxide breakdown field. Oxides of 600 nm thickness were grown at room temperature by plasma enhanced oxidation.

Ellipsometry measurements (by SF Wang of Liverpool) revealed oxide thicknesses in the range 575 to 620 nm, with maximum variation across a substrate of 5%, and refractive index 1.48 ± 0.03. High frequency CV measurements, Fig 4.6 yielded the result that the doping concentration in the epilayers was less than the nominal 10^{16} cm^{-3}, ranging from 1.5 to 4.6 10^{15} cm^{-3}. The control substrate had doping 1.5 x 10^{16} cm^{-3}. Quasi static C-V revealed mid-gap interface state densities typically 8 x 10^{11} cm^{-2}. These relatively high values were also found on the control. More problematic were the oxide breakdown fields which range from 4 MV to only 0.2 MV cm^{-1} which corresponds to V_g of 12V. This breakdown could be catastrophic,

P type Si

dVdT = 0.203 V/S
Cox = 395.53 pF
Cmin = 111.357 pF
depth = 686 A
dia = 1 mm
area = 7.86e-3 cm^2
Na = 3.28e15 cm^{-3}
Nf = 2.45e12 cm^{-2}
Cfb = 234.71 pF
Vfb = -8.815 V
Cmid = 150.43 pF
Vmid = -6.047 V



113.25 1.0mm A
14:14 29-Aug-90 113.25a

Fig. 4.6 Typical high frequency capacitance voltage plot from an MBE grown Si epilayer.

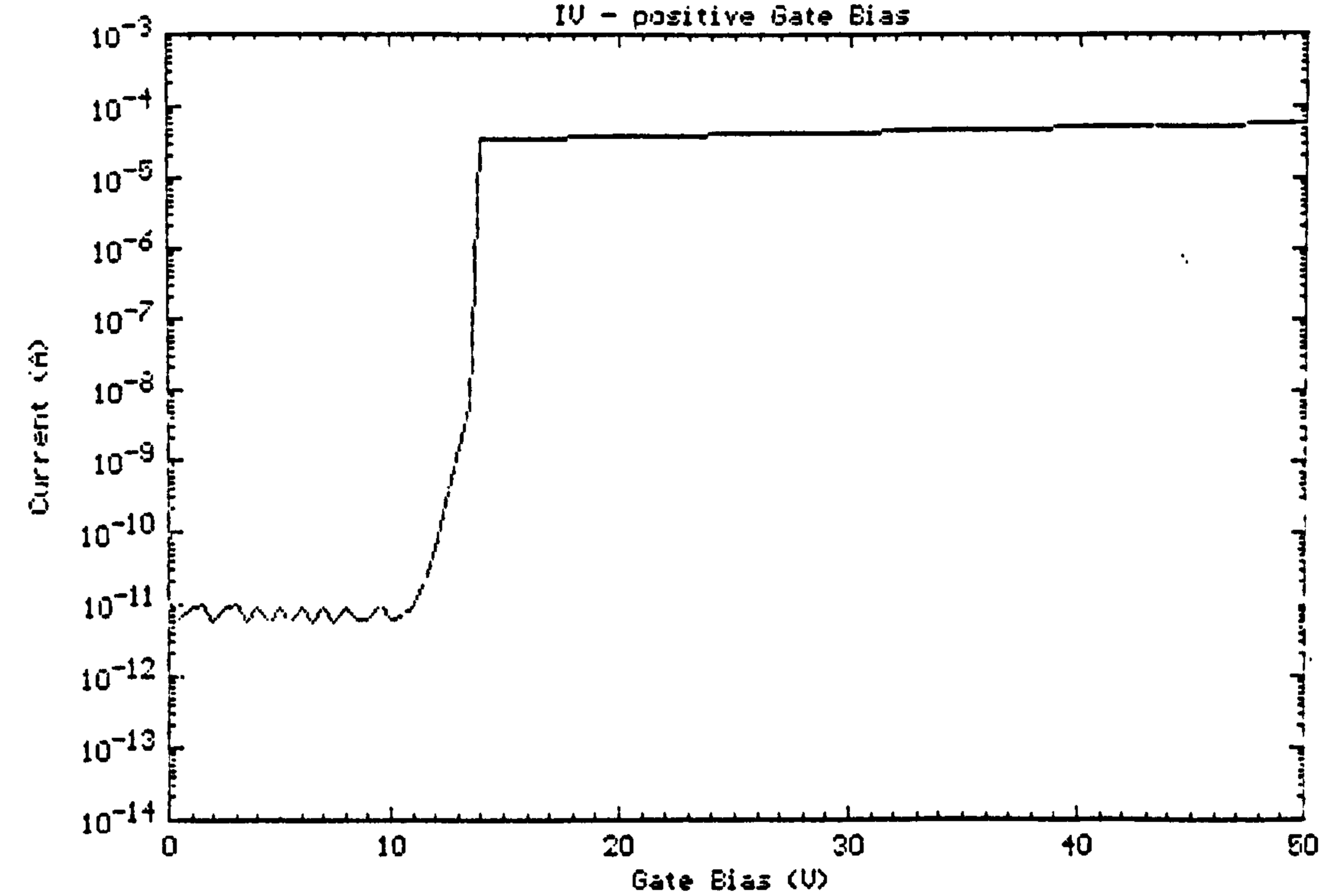
Fig. 4.7 (a) and (b). This represents a considerable restriction on depth profiling. The relatively poor oxide quality is considerable worse than any previously grown at Liverpool and was similar on the control. It is thought probable that this resulted from surfacial contamination introduced during transport of the epilayers. Unfortunately, no clean room chemical clean facilities were available at Liverpool.

The capacitors were pulsed from inversion into deep depletion with gate pulses of between 4 and 15V. The observed transients fell into two categories. These are illustrated in Fig 4.8 (a) and (b). Category (a) behaviour was seen from all capacitors formed in the epilayers grown at 500C and 750C. The other samples yielded Category (b) type behaviour. A general description and assessment of this behaviour is useful before attempting quantitative analysis. Category (a) behaviour is a very rapid C-t response, where the initial capacitance is anomously low (for these N_a and V_g), and in a form that yields non-linear Zerbst plots. Category (b) behaviour can be described as the expected MOS C-t response for low voltage pulses, but exhibits a distinct two-stage response for higher voltage pulses, the first stage being a very rapid capacitance relaxation.

What is the cause of the two stage (Category b) behaviour? There are three possibilities; field enhanced generation, depletion into the interface and oxide breakdown. This first, field enhanced generation is the effect whereby the applied E field distorts the potential well at a MGS, Fig 4.9 (a), and lowers the barrier to thermionic emission. This can increase the generation rate, G, to a field dependent value $G(E) = G_0 e^{q\Delta\phi/kT}$ and, so $\tau_g(E) = \tau_{g0} e^{q\Delta\phi/kT}$ **4.33**

It follows that $\ln \tau_g(E) = \ln \tau_{g0} + \frac{\sqrt{E}}{kT}\sqrt{\frac{q^3}{\pi\varepsilon_0\varepsilon_r}}$. **4.34**

This process can be confirmed, therefore, by plotting $\ln\tau_g$ versus $\sqrt{E}$. No useful Zerbst plot could be obtained from the two stage behaviour, but, by inspection of the



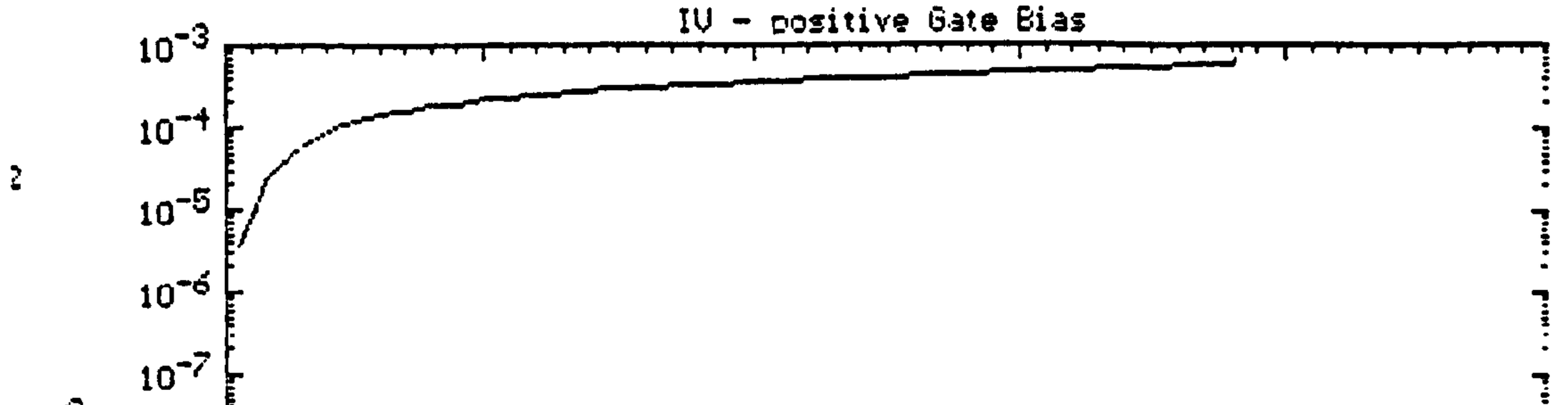


Fig. 4.7 Current versus gate voltage for oxides grown in the V80 Si study. (a) shows breakdown at 12V, (b) the characteristics obtained after catastrophic breakdown.

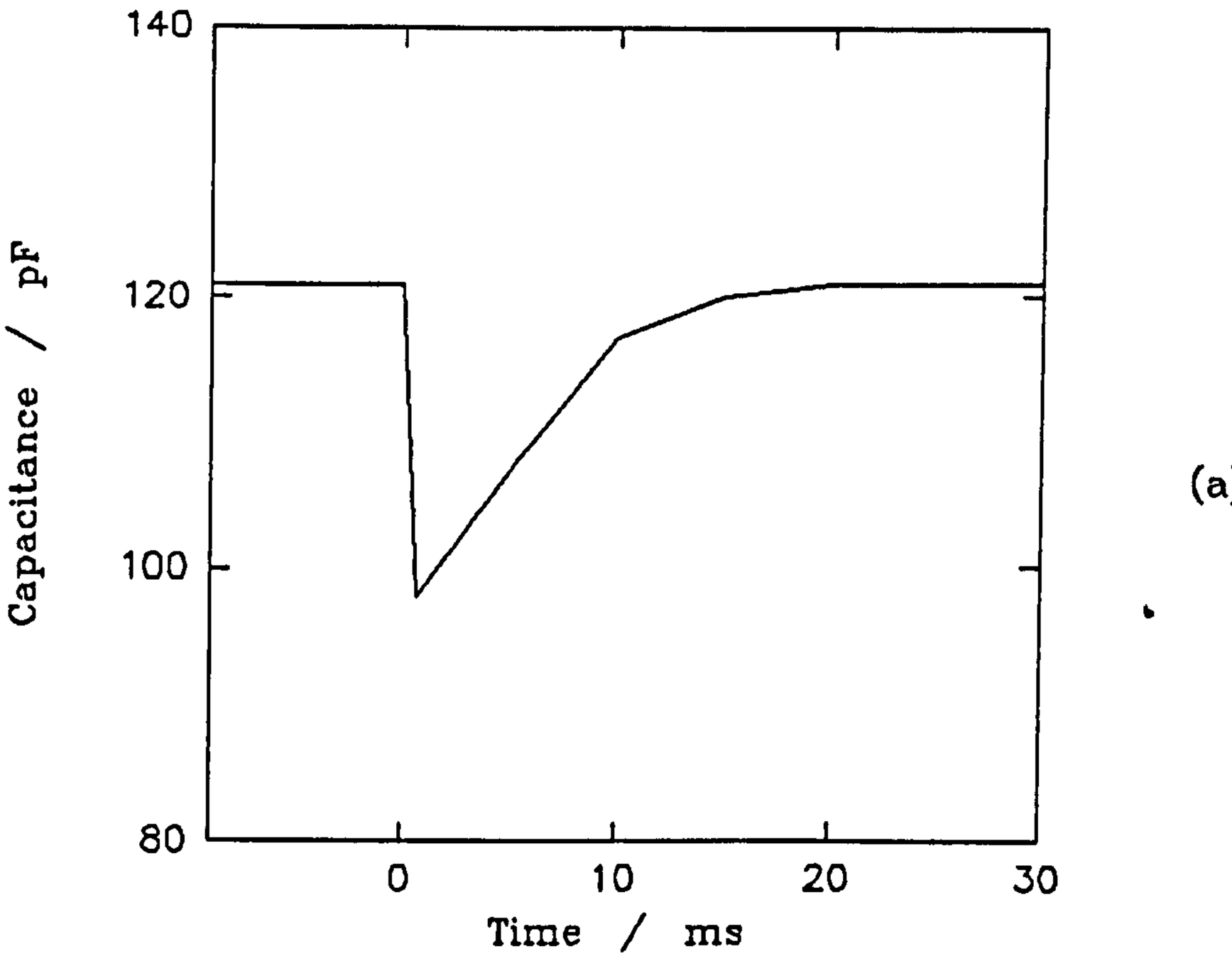


Fig. 4.8 Representative capacitance transients from the V80 grown Si study. Note the two stage behaviour shown in (b).

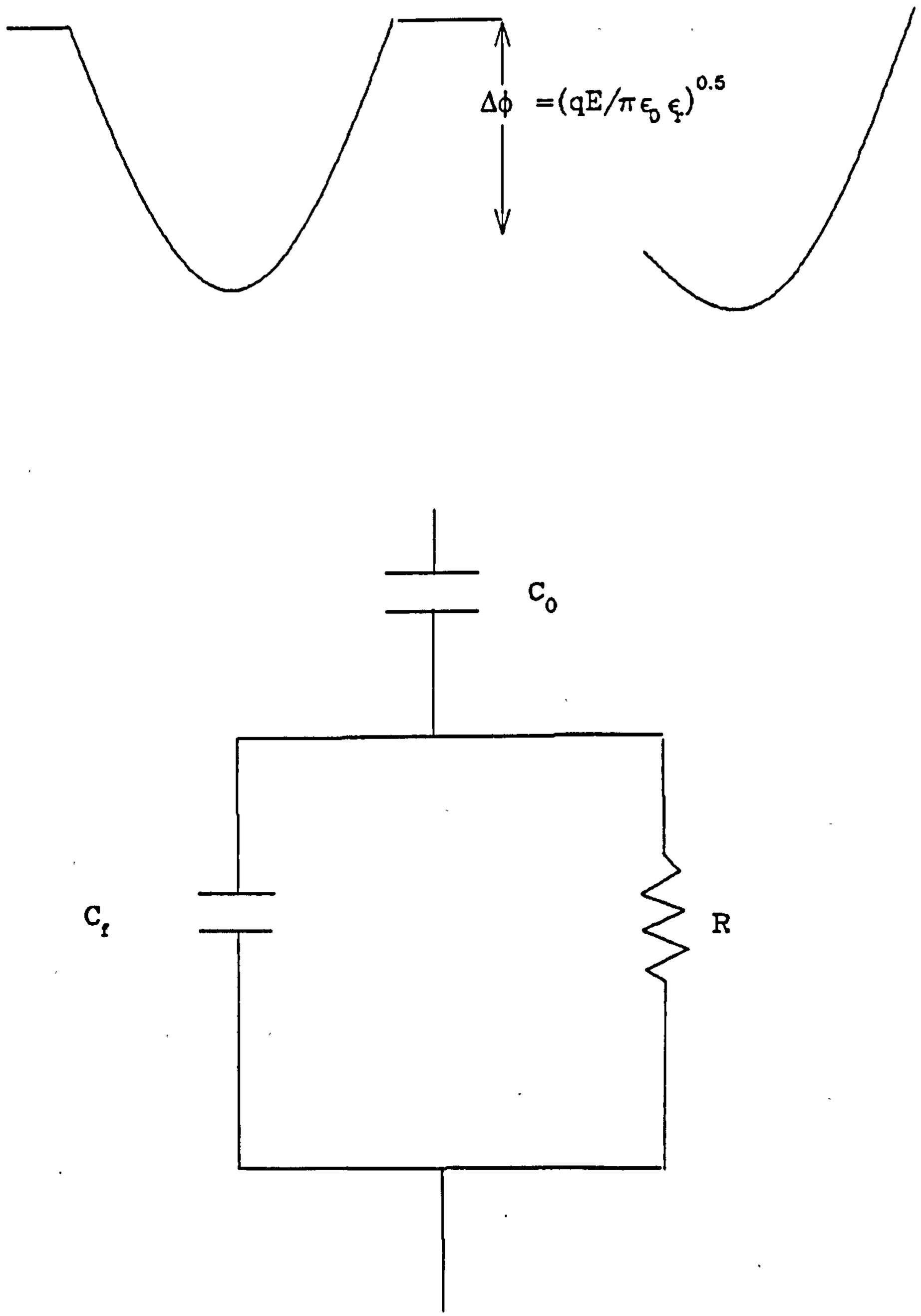


Fig. 4.9 (a) Barrier lowering at a midgap state due to an applied E field.
(b) Equivalent circuit for an MOS capacitor in strong inversion.

C-t, it is clear that this dependence is not dominant, so field enhanced generation cannot be the sole cause of the effect.

If the depletion width were to extend into the epilayer substrate interface (and the associated B spike), we might expect a low lifetime. This would tend to offset the increase in transient time resulting from the higher doping concentration. The initial rapid relaxation of depletion width seen in this study would imply a very much lower lifetime at the interface than in the bulk - at least three orders of magnitude lower. This is not inconceivable. However, increasing the pulse width further would result in a small decrease only of C, because in a high doped region only a small change in depletion width would be required to compensate the increase in gate charge. Further, the depletion width is 1.2 μm at 10V deep depletion, according to theory, which is still less than the epilayer thickness. Therefore this explanation is considered improbable. The most likely explanation is oxide breakdown.

The equivalent circuit of the MOS capacitor in strong inversion is shown in Fig 4.9 (b), where R is the parallel effect of resistances associated with minority carrier diffusion current and charge exchange at the interface. This circuit reduces to C(depletion) for $C_0 >> C$(depletion) and for R, $\omega \rightarrow \infty$. Oxide breakdown is equivalent to placing a further resistance in parallel with C_0, thereby lowering the total capacitance. The rapid restoration of capacitance is then due to transient RC discharge. The values of gate pulse at which this effect occurs, are consistent with the measured oxide breakdown voltages. This is also consistent with the rapid, and non-ideal, C-t response (Category (a) behaviour) obtained from the epilayers at 500 and 750C which exhibited consistently low oxide breakdown voltages. The evidence suggests that oxide breakdown is the cause of the observed behaviour.

This problem greatly restricted the depth profiling that could be achieved. More importantly it also restricted the number of capacitors from which linear Zerbst

plots could be obtained and therefore, limited the statistical significance of the results. For the case of two epilayers no valid results could be obtained at all.

The lifetime measurements from C-t's without oxide breakdown (i.e. for $V_g <$ 10V) are shown in Fig 4.10. No dependence of τ_g on T_s can be confirmed from these results for T_s in the range 600 to 800C. The epilayers have τ_g of approximately 200 ns, which is one order of magnitude lower than the control substrate subject to the same post growth processing. This value (200 ns) is less than that generally considered acceptable for commercial Si devices, but is not orders of magnitude lower than the ungettered control substrate. The absence of T_s dependence is not that predicted by the previous DLTS study of V80 grown Si (Sidebotham et al 1988), which would predict a decrease of lifetime as T_s is lowered from 700C to 600C: such a decrease is not confirmed from this data. It is highly unfortunate that the layer grown at 500C yielded no valid result. Doubts remain regarding the statistical validity of these lifetime values, but these results indicate that previously obtained very low values ~10 ns are not inherent to MBE growth.

4.5.1.2 V90S Zerbst Measurements

Five Si epilayers were grown in the V90S system to a thickness of 2um at a growth rate of 0.3nms^{-1}, with T_s in the range 500 - 700 C. These layers were B doped at a nominal concentration of 2 x 10^{16} cm^{-3} and grown on 4" p^+ substrates. No ex-situ cleans were employed. Two p-type, 10 Ωcm Czochralski, back-damaged substrates were used as controls, one of which was subject to transport around the MBE system and to the in-situ pre-growth flux clean, but no Si was deposited. These epilayers and the two control substrates were then oxidised at 100C via RF plasma oxidation to produce a 500 nm oxide. The growth orders of both the Si and SiO_2 were randomised with respect to T_s and to each other (see Table 4.1).

T_s / C	500	550	600	650	700	Control A	Control B
Wafer I.D (24 series)	31	33	29	32	34	35 (loaded)	------
N_a /cm^{-3} x10^{16}	1.20	0.92	1.26	2.00	1.1	0.15	0.3
D_{it} /cm^{-2} x10^{11}	4.7	2.9	1.6	5.2	1.3	3.7	3.5
τ_g (mean) / μs	0.65	1.92	0.17	0.09	1.83	25.0	6.15
σ^{n-1}	0.29	1.16	0.11	0.05	0.38	5.3	2.62
s (mean) /cms^{-1}	8.3	17.4	14.3	5.2	4.6	2.3	1.3
Dislocation /cm^{-2}	<40	~80	~50	~50	~70		
s-pits / cm^{-2}	$1x10^4$	$5x10^3$	$8x10^3$	$7x10^4$	$3x10^4$		

Table 4.1 Parameters derived from MOS capacitor measurements and defect etching from Si grown in the V90S and control substrates.

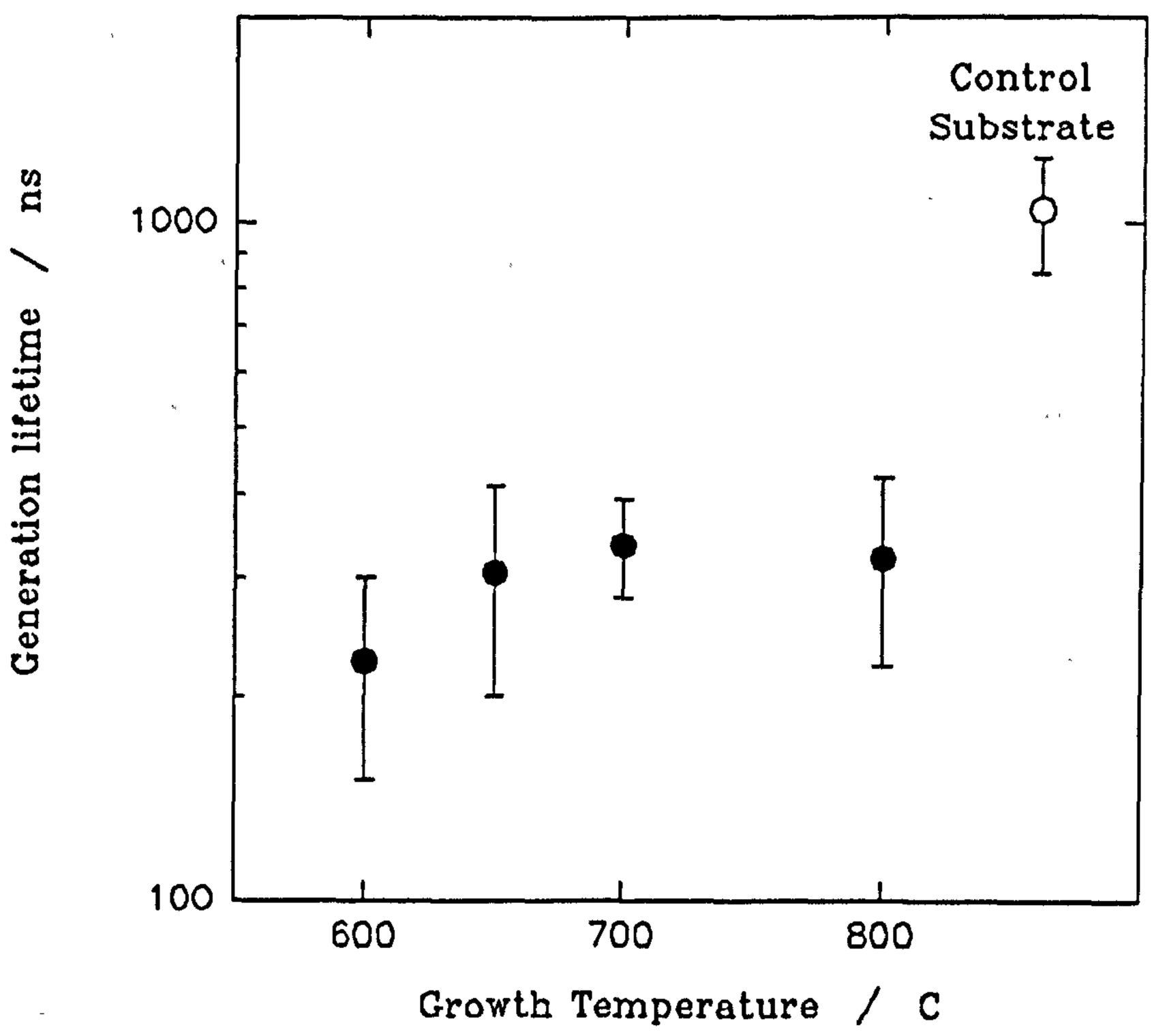


Fig. 4.10 Generation lifetimes as a function of growth temperature measured in V80 grown Si.

After oxidation, each sample was cleaved into quarters, one of which was annealed at 700 C for thirty minutes, another at 1000 C for fifteen seconds in an inert ambient. The first anneal represents the near-maximum anneal permissible for SiGe structures, the second is the implant activation anneal. Both anneals were carried out at the EMF to replicate anneal conditions during processing.

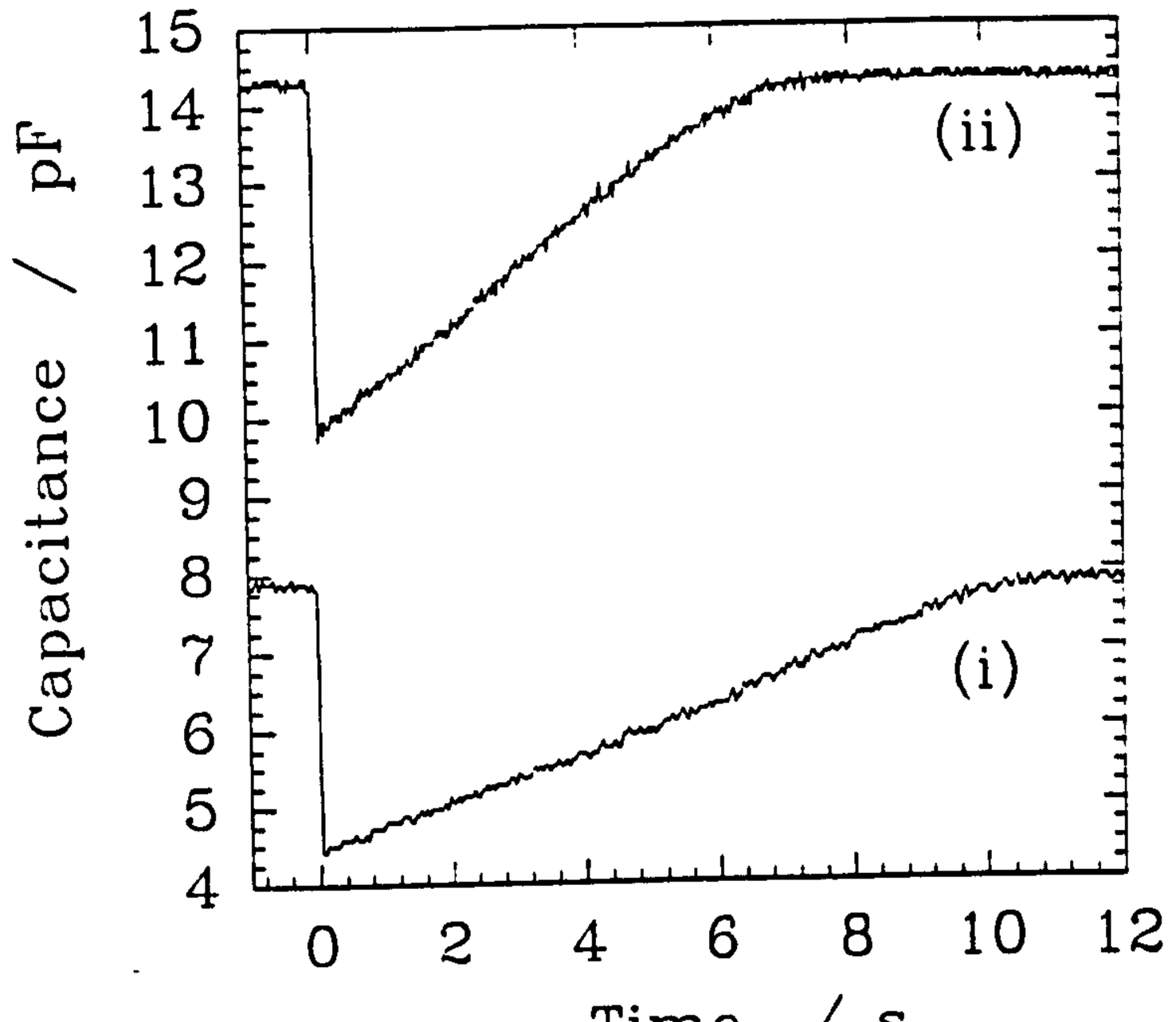
Subsequently Al contacts were formed, then CV and C-t data taken. After electrical analysis, a portion of each layer was stripped of oxide in HF and then Schimmel etched for defect reveal. The high frequency CV measurements confirmed doping levels of 1.5×10^{16} cm^{-3} $\pm$ 5×10^{15} cm^{-3} in the epilayers: quasi static CV revealed D_{it} of 2×10^{11} cm^{-2} $\pm$ 1×10^{11} cm^{-2}. By comparison, these act as dual frequency CV measurements and confirmed the absence of series resistance effects (these effects can otherwise lead to appreciable error). Oxide breakdown occurred at E fields in the range 4 to 8 MV cm^{-1}.

The C-t measurements were taken by pulsing from inversion ($V_g = 4V$) to deep depletion. This, rather than pulsing from accumulation, was designed to mitigate against the variation of surface recombination velocity during the transient. This technique allows surface equilibrium to be reached in inversion, but requires previous measurement of C_0. Work by Schroder and Nathansen, 1970, highlights the possible errors from lateral depletion about the contact. This is expected to be negligible for this case with contact diameter much greater than the depletion width.

4.5.1.3 Results

The Zerbst plots all yielded good straight lines from which τ_g and s could be extracted. Examples of C-t's are shown in Fig 4.11 (a) and the resulting Zerbst plots in Fig 4.11(b). There are two regions of non-linearity: at high $C_f/C-1$ (t=0) due to surface generation, and at low $C_f/C-1$ as equilibrium is approached and the condition n,

(a)



(b)

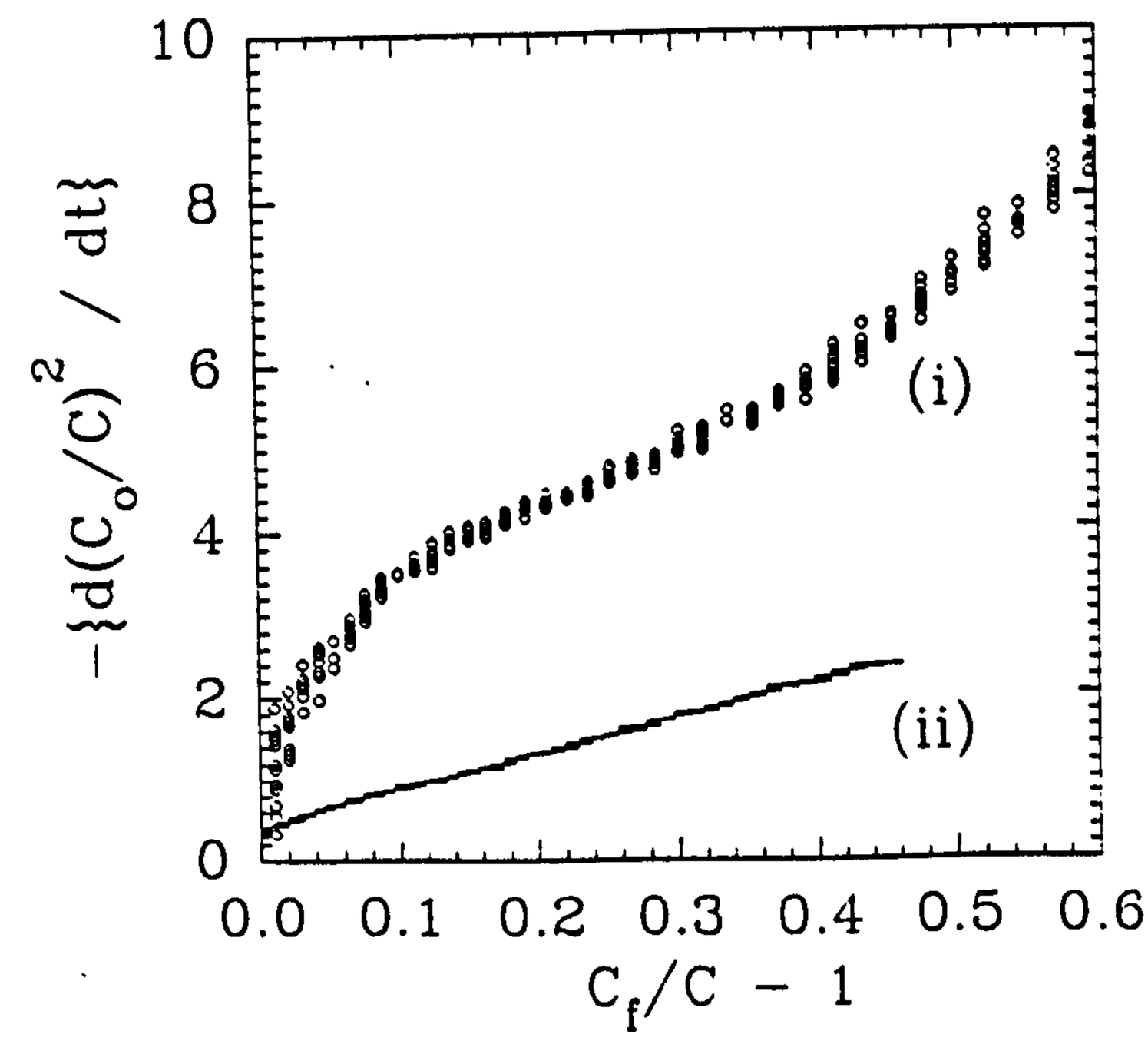


Fig. 4.11 (a) Capacitance transients and (b) the resultant Zerbst plots from a control substrate (i) and V90S grown Si (ii).

$p < n_i$ is not met. These regions of non-linearity are reported by many authors (e.g. Zerbst, 1966). The surface generation induced non-linearity was seen in approximately one fifth of the transients taken and was not associated with a particular epilayer, but rather, specific capacitors; indeed, they also occurred on the controls. Values of s were below 50 cms^{-1}: there was no correlation between high s and low τ_g on any given sample. This suggests that variation of s during the measurement was negligible.

The mean lifetime is shown as a function of T_s in Fig 4.12, where the bars represent the range of values obtained from each sample, not the error in the measurement. Repeated measurements on a specific capacitor yielded 10% variation about the mean value, which can be taken as the measurement and analysis error. Depth profiling for τ_g, by increasing the pulse width, ΔV_g, was limited by oxide breakdown. However, ΔV_g in the range 2V to 9V revealed approximately constant τ_g, although its variation of 25% was greater than that obtained from repeated measurements at fixed ΔV_g. The large range of τ_g represented by the bars result from variations from capacitor to capacitor across each sample. Again, no consistent lateral trend was observed, so this, presumably, is the result of an inhomogeneous distribution of defects in the epilayers. (Note that the controls yielded more consistent values of τ_g.)

The results taken from the annealed samples are shown in Fig 4.13. Within the observed variation across any sample, we must conclude that the τ_g is unaltered by either anneal.

The most striking result is the strong dependence of τ_g on T_s, a minimum of ~ 90 ns being found at T_s = 650C, which is at least an order of magnitude lower than the highest values. This is unexpected and could merely reflect a random variation of material quality or oxide induced contamination. To test this possibility, two further sets of three epilayers were grown - one set in the same growth series as the original data, the other in a subsequent series. Each set consisted 2 μm epilayers, B doped at

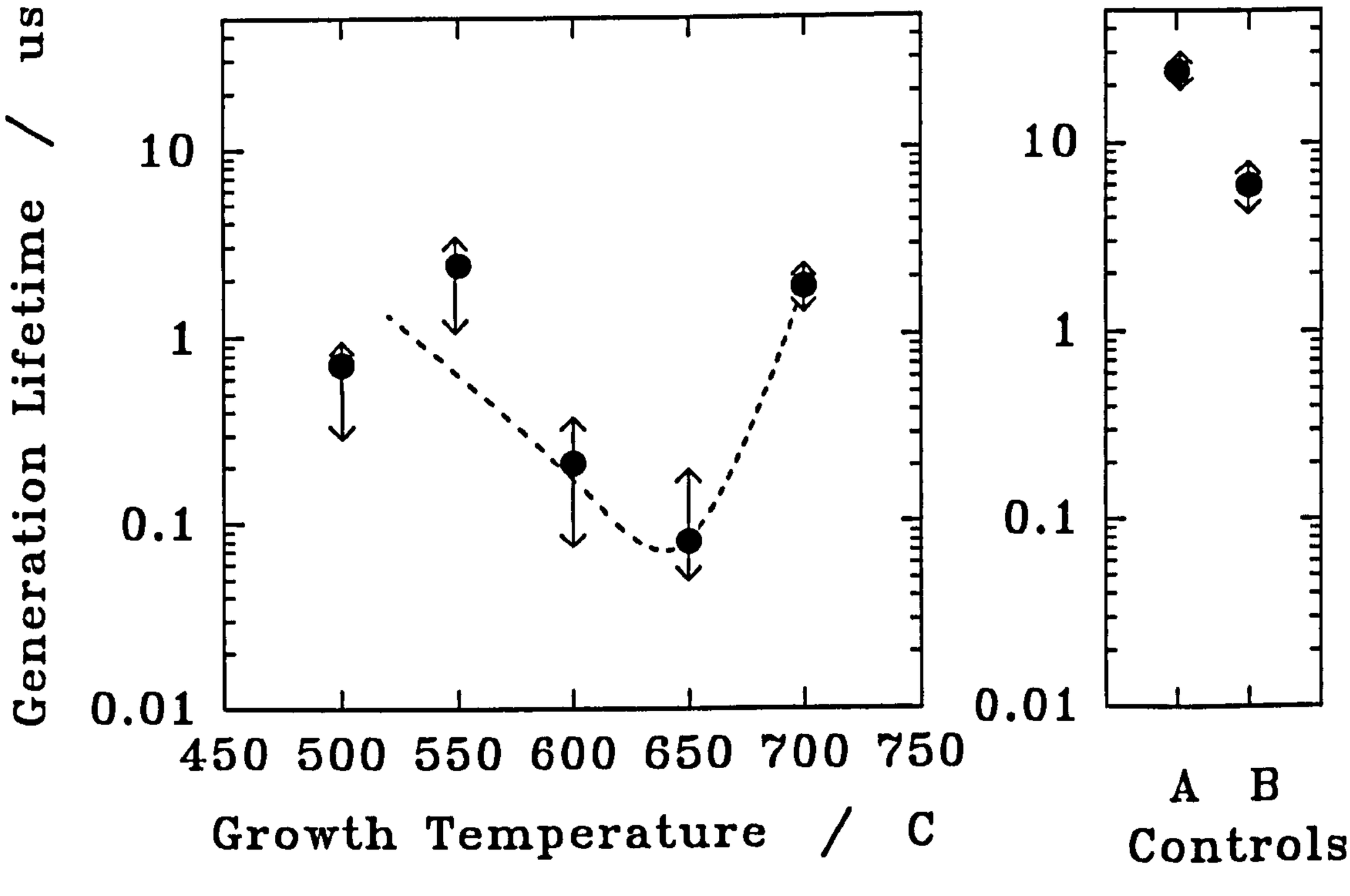


Fig. 4.12 Generation lifetimes measured in V90S grown Si as a function of growth temperature.

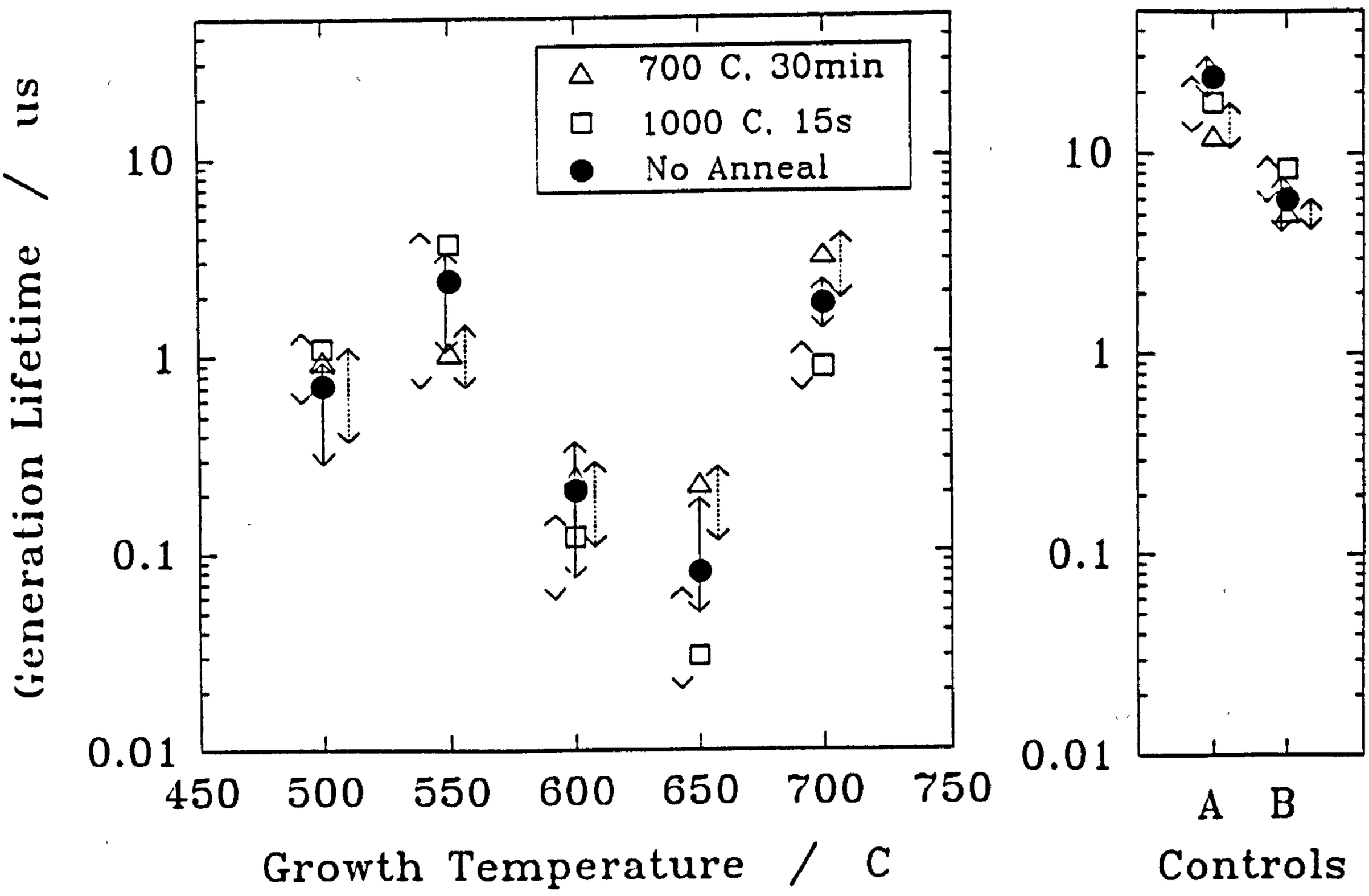


Fig. 4.13 The effect of two anneals on the generation lifetimes in V90S Si. (The unannealed data is repeated from Fig. 4.12 for comparison.)

10^{16} cm^{-3} as before, with T_s equal 530, 600 and 700C. To these were added two further control substrates. Oxidation was performed on all eight samples consecutively, but in a random order with respect to T_s. No anneals were employed. The results are shown in Fig 4.14. We believe that the reproducibility of results from these three sample sets confirm that τ_g is a strong function of T_s - with order of magnitude variations - and has a minimum value at T_s of 600 to 650C (at a growth rate of 0.3 nms^{-1}).

Defect etch results are also shown in Table 4.1. Very little variation between the samples was observed. Dislocation densities are low, less than 50 cm^{-2}, and the dominant defects are s pits in the range of 10^3 to 10^4 cm^{-2}. No stacking faults were identified, these being normally characteristic of high temperature processing e.g. steam oxidation.

4.5.1.4 Discussion

The highest lifetimes found in this study, at T_s=550 and T_s=700C, are of the same order as those found in control substrates, although between two to five times lower. The lifetime found in the control substrates indicates that no significant degradation of lifetime is introduced through substrate loading and pre-clean procedures. The T_s dependence is strong and is not equivalent to that observed from Zerbst studies on V80 grown Si (nor that expected from DLTS studies of V80 grown Si). The 1000 C implant activation anneal has not altered this dependence, nor has the 700 C 30 minute anneal. This latter anneal represents the thermal budget for 0.5 um growth at the highest T_s in this study. This Zerbst study sampled the material quality of the Si epilayer from 0.3 μm to 1.0 μm below the surface. If the T_s dependence were the result of solid state diffusion effects, then, by replicating the diffusion during growth of the sample grown at T_s=700C, this anneal would tend to smear out the

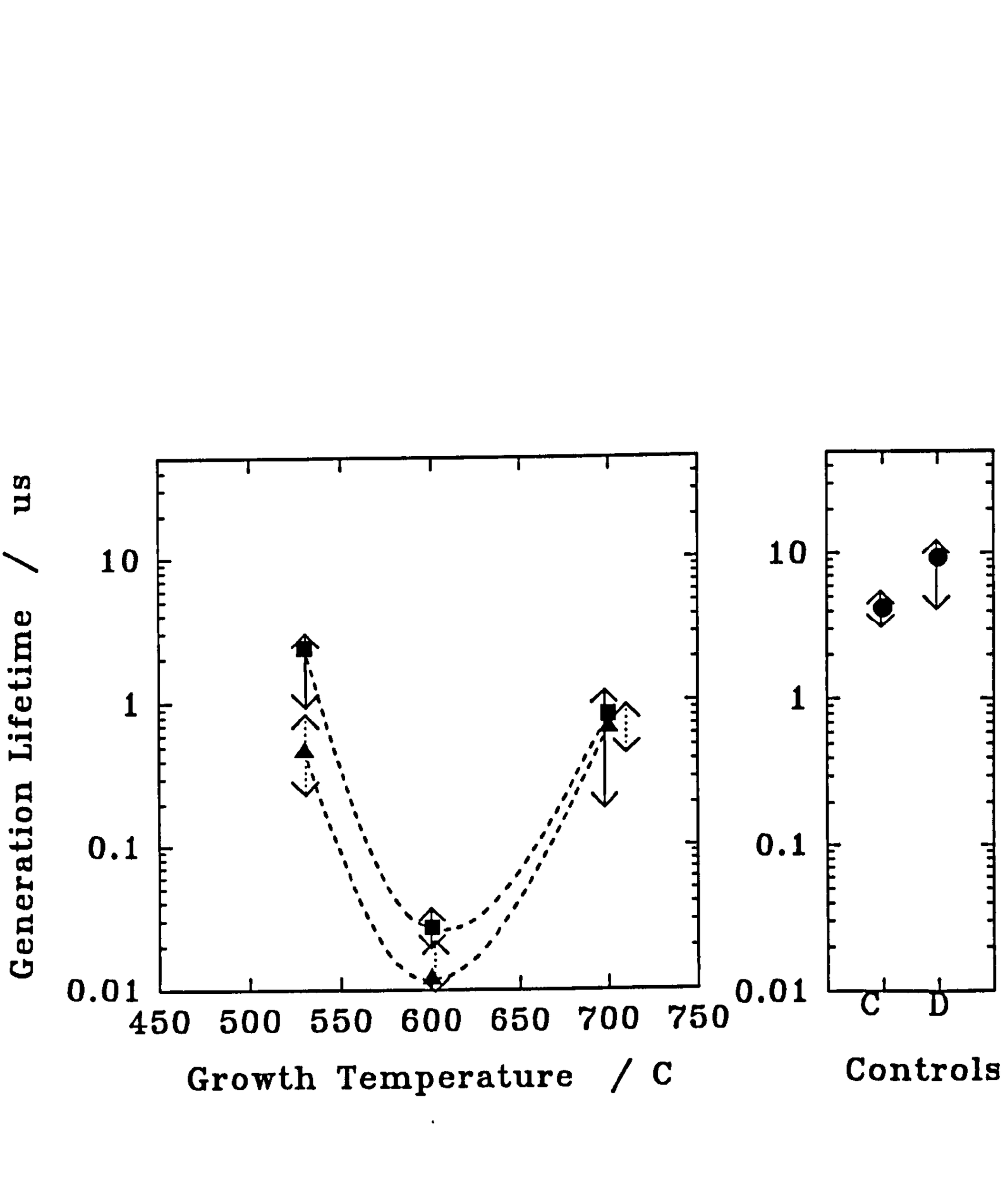


Fig. 4.14 Lifetime results from two supplementary, V90S grown sample sets. Squares represent layers grown in the 24 series, triangles those grown in the 28 series.

relationship between growth temperature and τ_g. Therefore, this result suggests that the T_s dependence is a growth related effect. The defect concentrations are approximately constant with T_s, which suggests that the τ_g dependence is not related entirely to crystalline defects.

It is possible that the T_s dependence is metal related, particularly since the dominant defects are s-pits. By comparison with the T_s dependence of B solid solubility concentrations (Chapter 2), it is clear that impurity incorporation/precipitation at the growth surface can have a complex dependence on T_s. However, it is noted that B substitutional incorporation is a minimum at mid range T_s. If metal incorporation had a similar minimum, it would imply that the lifetime is dominated by metals on interstitial sites, or in complexes, rather than on substitutional sites. This would be surprising. However, B is smaller than Si, many metals are larger, this may be relevant to the incorporation kinetics as suggested by Pindoria et al, 1991. This speculation pre-supposes metals as the dominant RGC and that the rate of arrival of metal at the growth surface is constant. Therefore, in the absence of further corroborative evidence, it must be concluded that the cause of the T_s dependence remains unknown.

Stop Press

A broad feature at mid gap energies has been observed (by Brighten, 1993, using boxcar averaging DLTS), in Si grown by MBE in the V90S at a similar time as the Zerbst samples. This state, by virtue of its position in the band gap, could be considered the dominant RGC in a study of τ_g. The behaviour of this state is complex and no capture cross section could be found, so, no lifetime can be inferred. Although no firm attribution is made, it is suggested in this study that this might be a dislocation related defect. However, it is difficult to reconcile this with the low dislocation concentration observed in the Zerbst study. In order to gain further evidence regarding

the T_s dependence, the supplementary Zerbst sample set of three epilayers (taken from the same growth series as the original V90S Zerbst set), was prepared for DLTS analysis. The Al and oxide were stripped in HF, the samples were then RCA cleaned and TiAl Schottky contacts prepared, as described in Chapter 3. Despite four attempts, no adequate Schottky contact could be produced on the sample grown at T_s 600C, i.e. τ_g = 30 ns. The IV characteristics of the three samples were shown in Chapter 3. This is further conformation that problems encountered in forming Schottky contacts, to MBE Si, are primarily associated with material quality. This limits the information that can be gained from the DLTS analysis regarding the T_s dependence. The DLTS analysis is being undertaken for the author by Jan Peters at Bio Rad using a digital DL4000 Fourier Transform spectrometer. Unfortunately, this study was not complete at time of writing, but preliminary measurements indicate that the total MGS concentration is an order of magnitude higher in Si grown at T_s=530 C than at T_s=700 C (which is roughly in accord with previous work). The apparent contradiction between this result and the Zerbst study, however, may yet be resolved. The resolution and sensitivity of this DLTS system are higher than of boxcar averaging systems and preliminary results indicate there may be at least three separate MGS, of similar concentrations, at mid gap energies. An explanation of the discrepancy between the τ_g versus T_s dependence and DLTS studies may be contained within this observation. It is worth reiterating that the effect of a MGS in RG processes is not only exponentially related to E_t , but to the capture cross-section, so even MGS present at similar concentrations can make very different contributions to τ_g.

4.5.2 pn Diode Measurements

Pn diode reverse capacitance and current measurements have been made on V80 Si epilayers grown at T_s=700 C and on V90S epilayers grown at a range of T_s. These latter structures differ from those of the Zerbst analysis sample set in important ways. Firstly, the epilayers are grown on 3" substrates; thus, an RCA clean has been employed prior to growth. Secondly, the first 200 nm of Si is Sb doped, to diminish the effect of the interface; thus, these n type regions were grown at 750 C, before T_s was ramped to the required value during growth of the B doped region.

The first study, on V80 material, yielded evidence of strong surface effects, described in section 3.6.5. A simple analysis provides a lower limit of τ_g even in the absence of a linear $I_r \propto V_r^{1/2}$ dependence. By assuming that the reverse current (below 18v, which is the onset of a large saturated current) is entirely due to generation current, a comparison with theory yields a lower limit of $\tau_g \approx 10$ μs. This value is significantly higher than that measured by Zerbst analysis, or that extrapolated from previous DLTS studies on V80 Si (predicted τ_g ~1μs). It is possible that the processing has had a gettering effect. However, the surface effects prevent a thorough investigation.

The second study, on V90S material, consisted of IV and CV measurements on the structures shown in Fig 4.15. These were processed at the EMF using mask Eu 931, with final passivation provided by an ECR oxide. The diodes have been reversed when compared to the V80 structures i.e. the low doped side is the upper layer. This was considered necessary for recombination lifetime studies to be discussed later. Representative $1/C^2$-V, N_a-W and I versus W plots are shown in Figs 4.16 and 4.17. The linear dependence is typical and indicates that surface effects are negligible. The lifetimes were extracted from these data using the quasi Zerbst plot of Fig 4.17. The generation lifetime is shown as a function of T_s in Fig 4.18. The T_s dependence seen

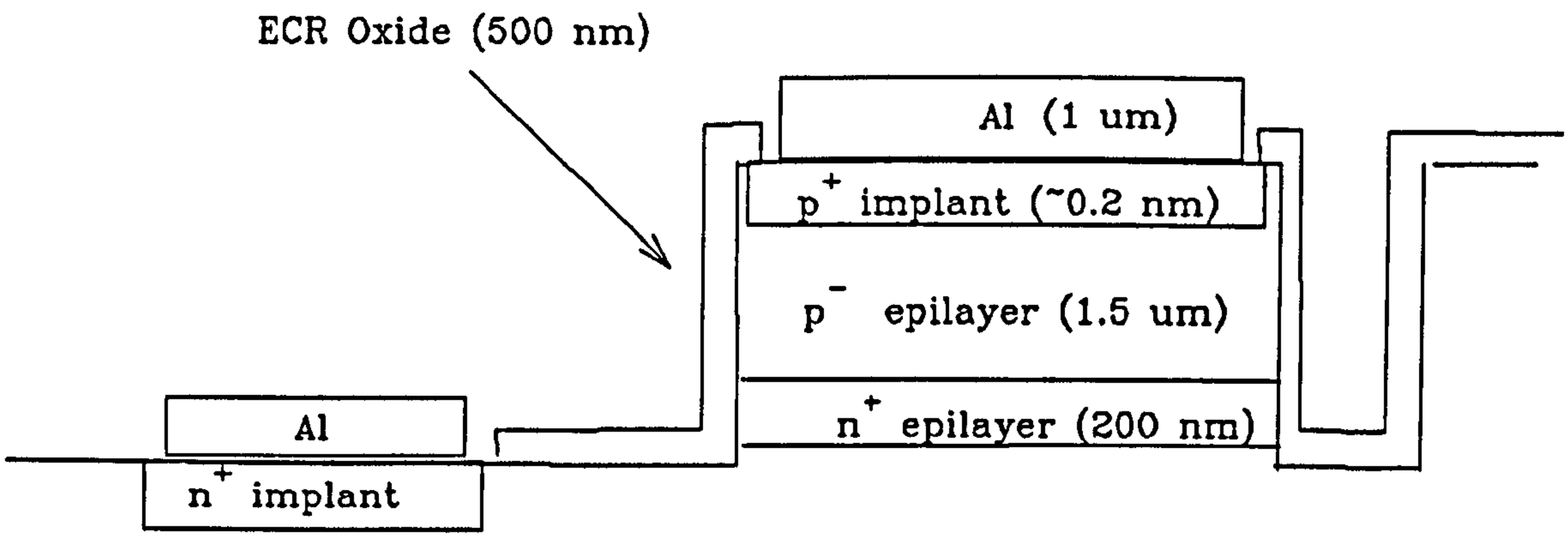


Fig. 4.15 Schematic of the pn diode used to assess T_g dependence of material quality for processed MBE Si.

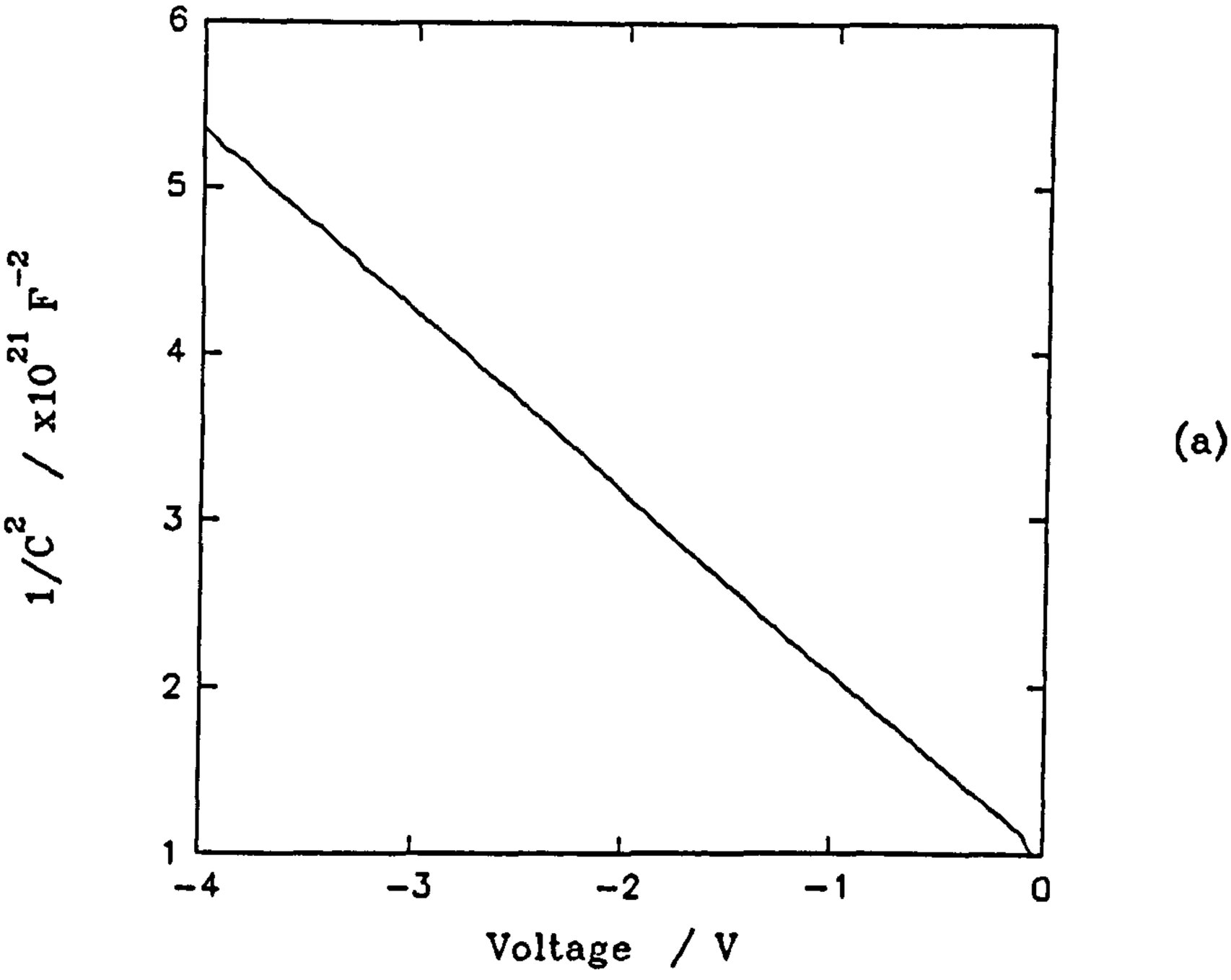


Fig. 4.16 (a) 1/C2 versus reverse bias plots obtained from the structures of Fig. 4.15 and (b) the resulting carrier concentration profile.

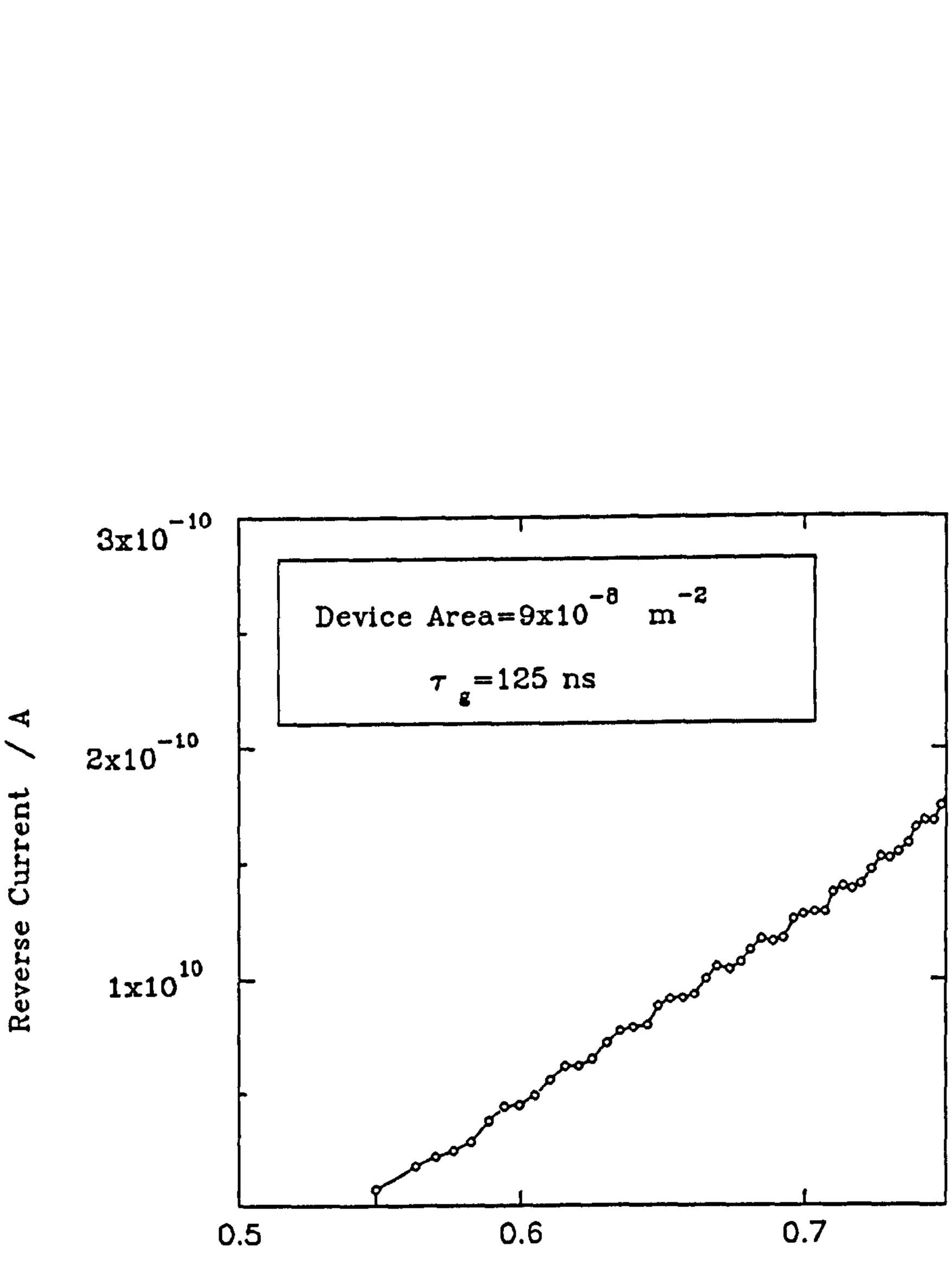


Fig. 4.17 Reverse current versus depletion width from a diode grown in the V90S system at T_s = 500 C.

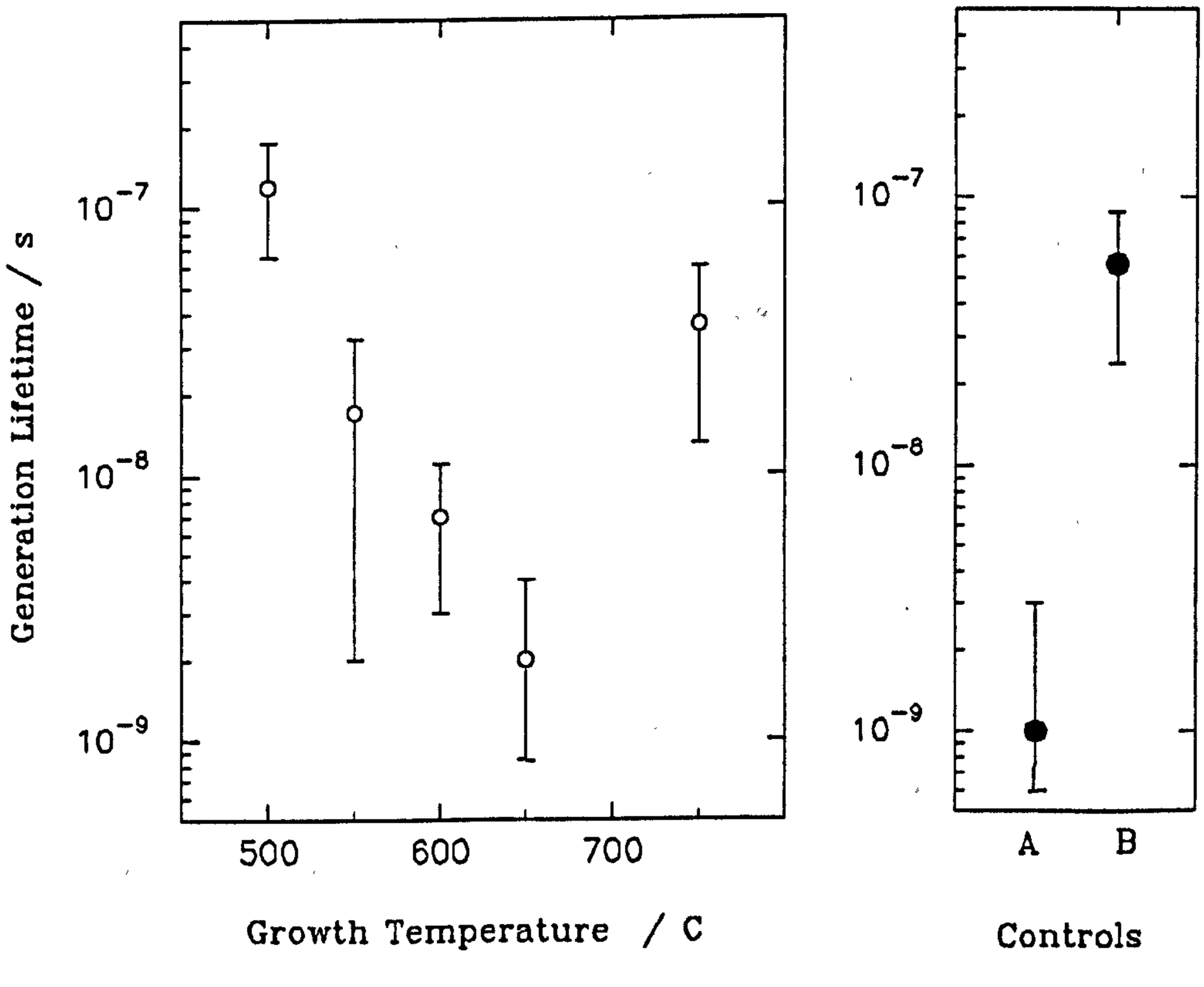


Fig. 4.18 Generation lifetimes extracted from V90S diode current and capacitance measurements.

dependence seen in the V90 Zerbst study, i.e. with minimum processing, is not seen in these samples and the lifetimes are generally lower. Either these epilayers exhibit a random variation of electrical Si quality or the substrate clean is of paramount importance, or, more likely, processing has introduced deep states. The control substrates also have widely varying, and generally low, τ_g, which supports the latter hypothesis.

These results do not, it is suggested, reflect the quality of MBE grown Si. This is further confirmation that material quality studies can be highly dependent on the history of the MBE grown structure. Evidence has been presented that the anneals used in these processing schedules do not significantly alter τ_g; however, this result will be critically dependent on furnace cleanliness. Continual monitoring of process steps, during processing of MBE test structures, will, of course, be as important as it is for commercial bulk Si applications. The pn diode is a relatively simple structure that may suitable for such monitoring, but also has greater significance for τ_r studies (see following Section).

It is worth noting that the values of τ_g derived from pn diode measurements assume, as is usual, that generation occurs throughout the full depletion width of the pn diode. Calzolari and Graffi, 1972, and then Braun and Grimmeiss, 1973, have shown that generation only occurs in a fraction of the depletion width. This fraction being dependent on the relative positions of the quasi Fermi level (thus bias voltage), and the mid gap state energy level. The true generation rate, therefore, is higher - and so τ_g lower - than that derived assuming uniform generation in the full width. This implies that IC is not a constant. However, I versus W plots from this study were generally linear and the application of this correction is not trivial, so the usual assumption is used. Zerbst analysis, in contrast, assumes generation only in the non-equilibrium depletion width ($W-W_f$). A comparison of lifetimes measured on the same sample by these two techniques would be both interesting and informative.

4.6 RECOMBINATION LIFETIME TECIINIQUES

As has been shown, τ_r is more sensitive to MGS than τ_g . It is also more difficult to determine in epitaxial layers. An excess of minority carriers will diffuse throughout a structure and, therefore, may be subject to four recombination rates; at the surface; at the epilayer-substrate interface; in the substrate and, of course, in the epilayer bulk. Any useful technique must either account for, or eliminate, these considerations. It must also be able to measure lifetimes that can vary from 100 ps to 1 ms. Measurements of τ_r have been made in the course of this study, but are believed to be erroneous (although lower limits may be obtained). This discussion will, therefore be brief and will focus upon the identification of a useful technique for future studies of τ_r.

There are a plethora of methods for finding τ_r in bulk material, which reflects the difficulty of so doing. These are well reviewed by Orton and Blood, 1990. Under low injection, it has been shown that the decay rate of excess carriers is exponential and is characterised by a diffusion length, $L=\sqrt{D\tau_r}$, where $D=\mu(kT/q)$. Note that, here, μ is the mobility of minority carriers, μ_{min}. Although there is some debate about the value of μ_{min} as compared to majority carrier mobilities, the differences are not large. A plot of L versus τ_r is shown in Fig. 4.19 for $N_a=10^{16}$ cm^{-3}, the doping used throughout this study. If the maximum routine MBE epilayer thickness is 2 μm, then excess carriers injected at the surface will partly recombine at the interface region or in the substrate for all $\tau_r>70$ ns (two diffusion lengths). This rules out many bulk, optically stimulated techniques, i.e. Haynes Schockley, and the measurement of τ_r proposed by Calzolari et al, 1977, or Wei and Woodbury, 1985, using MOS capacitor methods. This general argument also applies to the technique used by Higashi et al, but insufficient details are given for a full critical analysis of the results. The pn diode

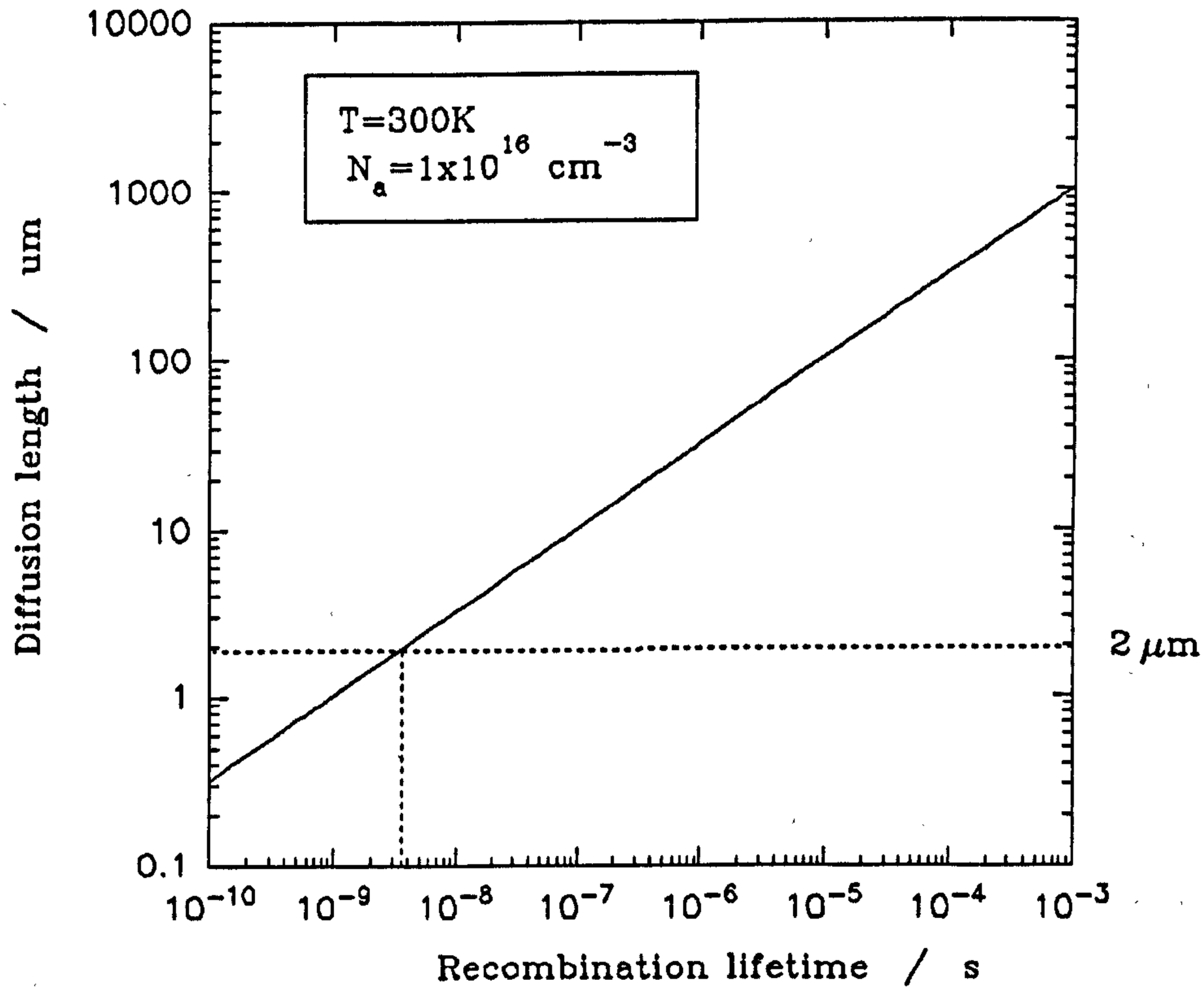


Fig.. 4.19 Diffusion length versus recombination lifetime for $N_a = 1x10^{16}$ cm^{-3} at room temperature.

was chosen as the basis for possible τ_r measurements in this work, due to its relative simplicity of growth and manufacture.

The base width, W, of a pn diode is defined as the neutral region in the low doped side from the depletion edge to the contact. A long base diode has W>L, short base, W<L. For short base diodes the excess minority carriers largely recombine at the ohmic contact, S=∞, so the time constant of carrier decay is the transit time across the base, $W^2/2D$.

Diode Switching

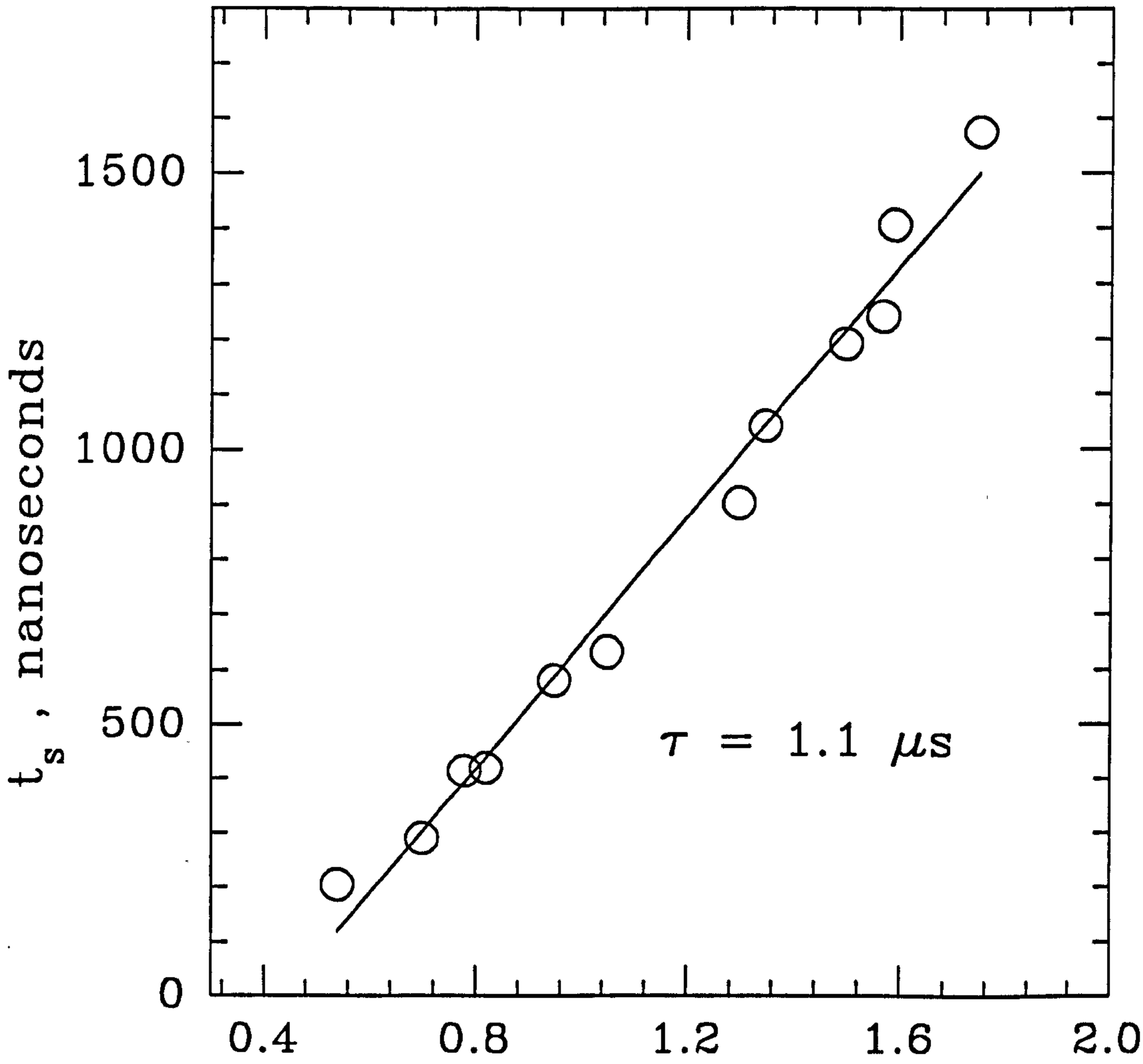
A diode in forward bias , with current I_f, stores charge in the base at the depletion edge. If the diode is switched to reverse bias, this charge must diffuse and recombine. The current, I_r, will remain constant during this time. Measurement of this storage time for various I_f/I_r allows τ_r to be found in long base diodes, e.g. Kingston, 1954.

The Kingston analysis was applied to the V80 grown diodes described earlier, Fig. 4.20, and a lifetime of 1.1 μs found. This corresponds to L≈ 30 μm, i.e. greater than the epilayer thickness. This result is, therefore, in part due to recombination in the substrate or interface. The fact that a proportion of electrons have reached this region implies a lower limit of τ_r~ 70 ns: no upper limit can be inferred. Note that the generation lifetime found from this sample set is ≥10 μs.

The analysis for switching in short base diodes was developed by Byczkowski and Madigan, 1957, but is cumbersome. A practical approximation is provided by Lewis, 1975,

$$\frac{1}{\tau_r} + \frac{1}{\tau_{transit}} = \frac{d}{dt}\left(\ln\left(1 + {}^{I_f}\!/\!_{I_r}\right)\right) \qquad 4.35$$

and estimates the errors to be small for $W/L > 10^{-3}$. Whilst the mathematical errors may be small, consideration of equation 4.35 reveals that for $\tau_r > \tau_{transit}$ (i.e. a short base diode), τ_r cannot be found in practice. Therefore, this technique is not useful.



$$\left[erf^{-1}\left(\frac{1}{1+I_r/I_f} \right) \right]^2$$

Fig. 4.20 The analysis of pn diode switching measurements from an MBE grown diode implying an erroneous epilayer recombination lifetime of ~1 μs.

As a matter of tactics it was decided to eliminate interface and substrate effects by reversing the diode, i.e. to grow the low doped side uppermost. This ensures that the experimental results are only complicated by the transit time, which can be calculated. This forces short diode behaviour for all but very low τ_r.

Diode Admittance

The frequency dependence of the forward bias admittance of a pn diode can be used to find τ_r in short base diodes. A series of publications have produced analyses of the applicability of this technique: these are summarised by Gonzalez and Neugroschel, 1984. This is an attractive approach since the experimental measurements can be simple.

Low frequency measurements, $\omega\tau_r << 1$, yield both S and τ_r provide that $WD/L^2 < S < D/W$. For $\tau_r << D/W$ τ_r may be found but S is not. Unfortunately, for high S, i.e. an ohmic contact, neither S nor τ_r can be found. High frequency measurements, $\omega\tau_r > 10$, can yield τ_r from measurements of $G(\omega)$ only, since the effective diffusion length decreases at high frequency, which renders the device long base. However, for the values of τ_r that might be expected this necessitates $\omega > 1$ GHz, so the technique is far from simple experimentally. The admittance method is, therefore, unattractive.

Modulation

A method for finding τ_r in short base epitaxial diodes has been proposed by Spirito and Cocorullo, 1985. A lateral short base diode is formed by implantation into the epilayer, Fig. 4.21. This diode is forward biased, as is the substrate with respect to the p^+ contact - the substrate bias repels minority carriers in the epilayer. If the substrate bias is modulated by a small ac signal, then the an additional modulated current is induced in the lateral diode. This modulation is due only to recombination in the epilayer because the dc diode current is a function of diode voltage only, which is held constant. Measurement of the various current components then yields τ_r. By

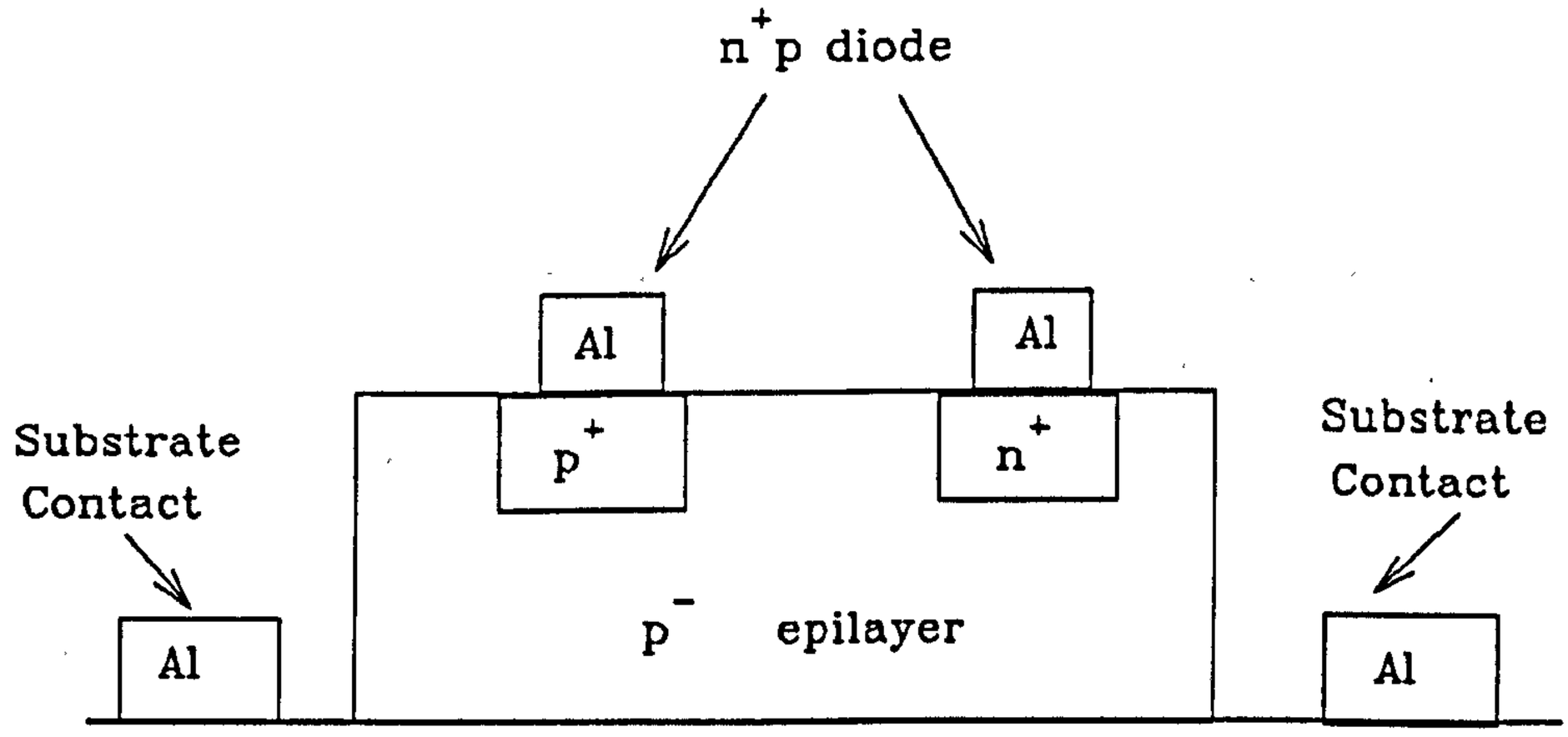


Fig. 4.21 Schematic of a structure proposed by Spirito and Cocorullo for recombination lifetime profiling.

varying the substrate bias, the width of the conductivity region of the lateral diode is altered so this technique allows limited τ_r profiling. A structure suitable for the application of this technique has been incorporated into mask Eu 931 and it is hoped to utilise this technique soon.

Unfortunately this method measures the high injection τ_r. A method for measuring τ_r under low injection conditions in the base of bipolar transistors has been proposed by Birrittella et al, 1979. This method requires the measurement of the low frequency output conductance and reverse transconductance only and includes an experimental check on the validity of the measurement. This may prove to be an effective technique and has been incorporated into mask Eu 931, with which the V90S grown pn diodes were procesesed. Unfortunately, time constraints have not allowed the author to assess this method, but the generation lifetimes measured in these samples indicated processing limited lifetimes, so this particular sample set is unlikely to yield particularly useful information regarding MBE material quality.

4.7 CONCLUSIONS

Generation lifetimes have been measured in MBE Si by two methods. Si from both V80 and V90S systems grown at a range of T_s has been assessed by an MOS capacitance transient method. By use of a low temperature oxide post growth processing temperature have been kept below the growth temperatures so the results are believed to reflect the quality of as grown Si. No T_s dependence can be confirmed from a study of V80 Si, which was limited by what are believed to be oxide breakdown effects. Generations lifetimes were found to be ~200 ns for T_s in the range 600 to 800 C. A further study of Si grown in a V90S system shows a complex but repeatable dependence of generation lifetime on T_s. The maximum values found of ~2μs compare to values of ~10 μs found in control substrates subject to the same processing, i.e.

ungettered. This T_s dependence is stable to post growth annealing and is therefore thought to be a growth related effect.

Values of generation lifetime extracted from pn diodes grown by MBE and processed at a VLSI facility are inconsistent and vary from 10 μs from V80 Si grown at T_s=700 C to a minimum of 1 ns in V90S Si grown at a range of T_s . Similar low, inconsistent values were obtained from control substrates. It is believed therefore that these values are defined by the post growth processing. Although the results are not irreconcilable with the previously observed T_s dependence, they cannot strictly be regarded as confirmation due to the process induced effects suggested by the controls. The only high temperature process, a 1000 C 15 second anneal does not inherently degrade the lifetimes.

Techniques for measuring the recombination lifetime in epitaxial layers have been discussed and useful methods identified. A lower limit of $\tau_r > 70$ ns has been measured in a processed MBE grown diode by a switching technique.

CHAPTER 5

MATERIAL STUDIES IN REMOTE DOPED SiGe CHANNEL 2D HOLE GASES

5.1 INTRODUCTION

Ge is, like Si, a group IV semiconductor that has a diamond unit cell in crystalline form. As such, it is an obvious candidate for alloy growth within a Si based technology. Thus, it is an attractive semiconductor alloy system for reasons of integration with current technology and the sophistication of near directly applicable processing techniques. This chapter discusses experimental studies of a 2D hole gas (2DHG) formed at a SiGe/Si heterojunction. This particular configuration might provide long hole coherence lengths in a relatively stable crystalline structure, via the technique of remote doping. This is a report of the effect of growth conditions on the low temperature mobility of the SiGe channel 2DHG. Low temperatures reduce phonon scattering and parallel conduction paths, which simplifies analysis of the 2DHG behaviour and provides information on material quality. The motivations for such research regarding fundamental solid state physics has, hopefully, been made clear in Chapter One. There is also an immediate application for this system as part of CMOS technology. A limitation of CMOS is that hole mobilities are lower than electron; thus, the p-type transistors must be larger than n-type in order to carry equal current. A SiGe heterojunction 2DHG with relatively high mobility might minimise this limitation, and is being actively sought. In CMOS, the 2DHG would, of course, be induced from a gate - enhancement mode, whereas B doping has been used as the hole source in this study. A description of the growth and origin of the quantum well in which the 2DHG is confined is followed by a review of the theory necessary for a discussion of the experimental results.

5.1.1 SiGe Heteroepitaxy

The primary difficulty to be overcome for heteroepitaxy is the lattice mismatch, which is 4% between Si and Ge since the lattice constants are Ge 0.566nm, Si 0.543nm. Si_xGe_{1-x} can be grown commensurately on Si and be strained to the substrate (or a relaxed alloy buffer) lattice constant or can relax to some extent. Relaxation occurs via the formation of misfit dislocations, which locally relieve strain at the interface, and are terminated by threading dislocations. Strained layers accommodate strain elastically by tetragonal distortion: depending on the relative lattice constants of substrate (or buffer) and overlayer the strain may be compressive or tensile. The tetragonal strain is given by $\varepsilon_{\perp} = f\frac{1+\nu}{1-\nu}$, where f is the mismatch obtained from Vegards' law and ν is Poissons ratio for the epilayer (ν_{Ge}=0.273, ν_{Si}=0.280 , Brantley 1973).

Strained growth occurs in layers up to an equilibrium critical thickness, t_c, which is a function of the lattice mismatch between the substrate and alloy. There is also a larger critical thickness, the metastable critical thickness, t_m, below which the epilayer is strained during growth, but can be relaxed by post growth annealing. Capping Si layers are generally known to preserve strain in metastable layers on subsequent annealing, but the effect of these caps on anneal-induced relaxation has not yet been quantified. Experimental values of t_c for the SiGe on Si system and theoretical predictions are shown in Fig 5.1. Low substrate temperatures are necessary for strained layer growth in order to retain localisation of misfit dislocations in relaxed layers and to ensure a sufficiently high yield stress for elastic distortion. Further, the difference in thermal expansion coefficients between substrate and epilayer is reduced, as is interdiffusion and Ge segregation, at low T_s. The onset of 3D alloy growth, or islanding, is shown as a function of T_s in fig. 5.2. Because of the aforementioned

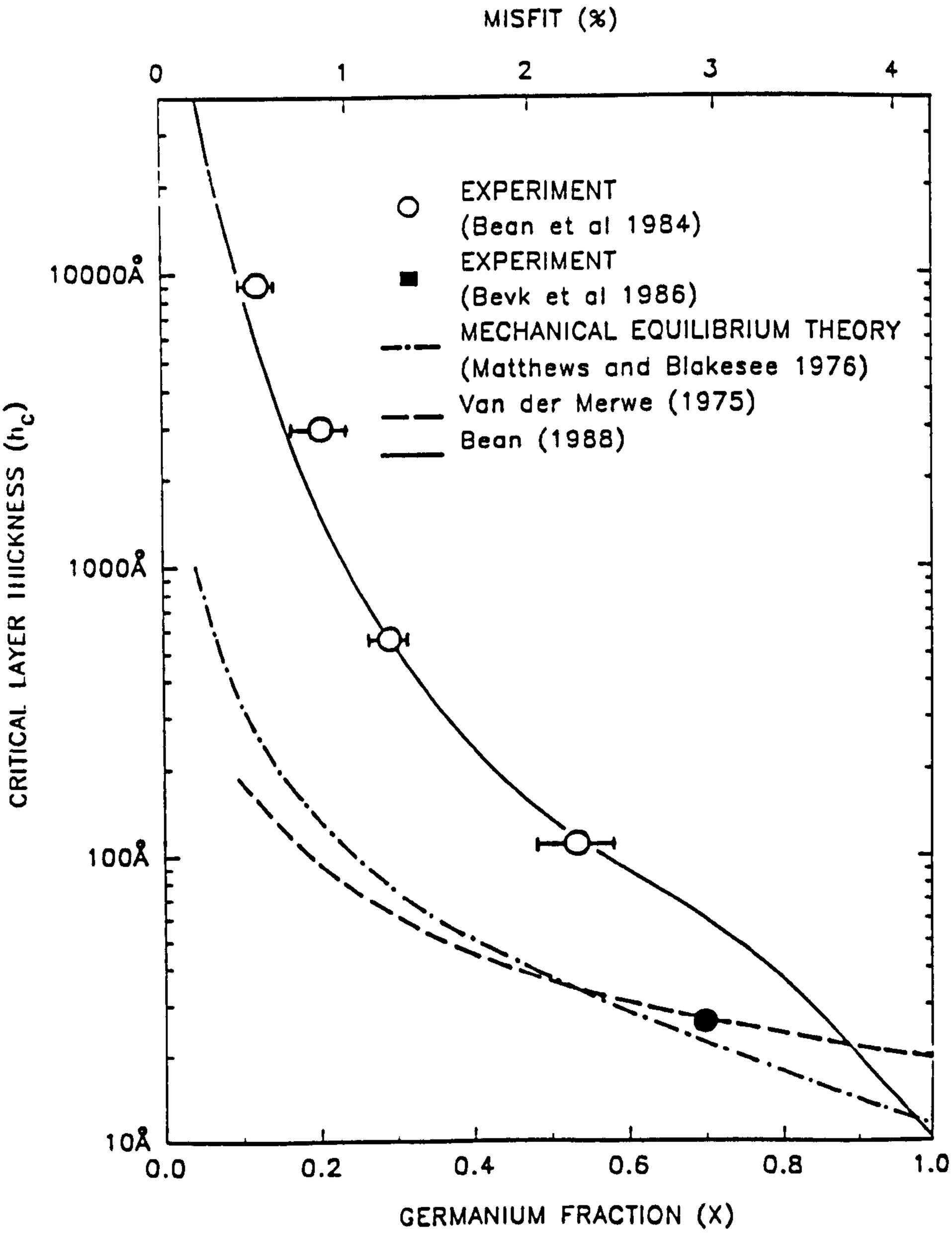


Fig. 5.1 Graph of critical thickness versus Ge fraction for $Si_{1-x}Ge_x$ layers grown by MBE. (After Iyer et al 1989)

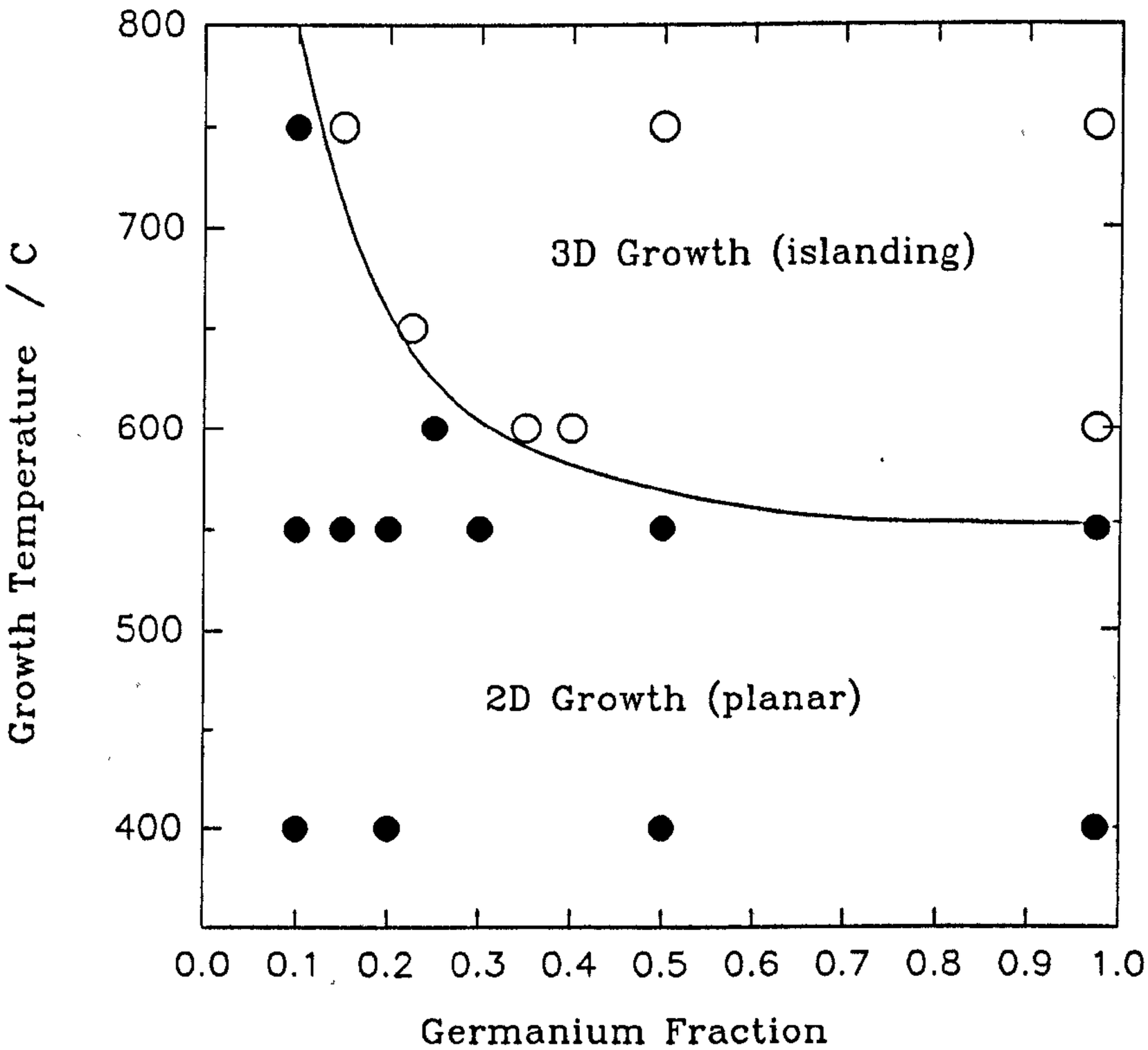


Fig. 5.2 $Si_{1-x}Ge_x$ growth morphology as a function of growth temperature, T_s. Filled circles represent data points, the curve is empirical.
(After Bean, 1988)

processes, research into the Si/SiGe system has concentrated upon minimising T_s for strained layer growth, which has, in certain respects, been a limitation, as will be shown in later sections.

Strained SiGe on Si growth was first demonstrated by Kasper 1975. The growth technology merely requires the addition of a Ge source to a Si MBE reactor. The vapor pressure and melting point of Ge are such that it can be evaporated from a Knudsen cell; however, in the V90S an e-beam evaporator similar to that of Si is used. There are two important differences between the Si and Ge sources used; firstly, that the Ge charge has a diameter of 2", compared to 4" for Si, secondly, that the Ge melt forms a convex meniscus while that of Si is concave. This will be of critical importance to this chapter.

5.1.2 Remote Doping

The concept of remote doping was introduced by Dingle et al 1978 as a means of increasing carrier mobilities. Carrier scattering in Si at high temperatures (>100K) is dominated by acoustic phonon interactions. At low temperatures, $4.2K < T < 70K$, ionised impurities provide the principle scattering mechanisms as described for 3D by either the Brookes-Herring or Conwell-Weisskopf theories (see for instance Chattopadhyay and Queisser, 1981). If the charge carriers can be spatially separated from the parent atom then mobilities could be expected to increase and, in fact, do. A similar situation can be achieved in the surface channel of Si MOS systems biased to strong inversion, where the carrier density locally exceeds the ionised impurity concentration. However, the reduction of Coulomb scattering can be far more efficient using heterojunctions. Remote doping requires a quantum well to act as a confining potential. An example is shown in Fig 5.3. A region of wide gap semiconductor is highly doped and separated from the quantum well formed at a heterojunction by a low

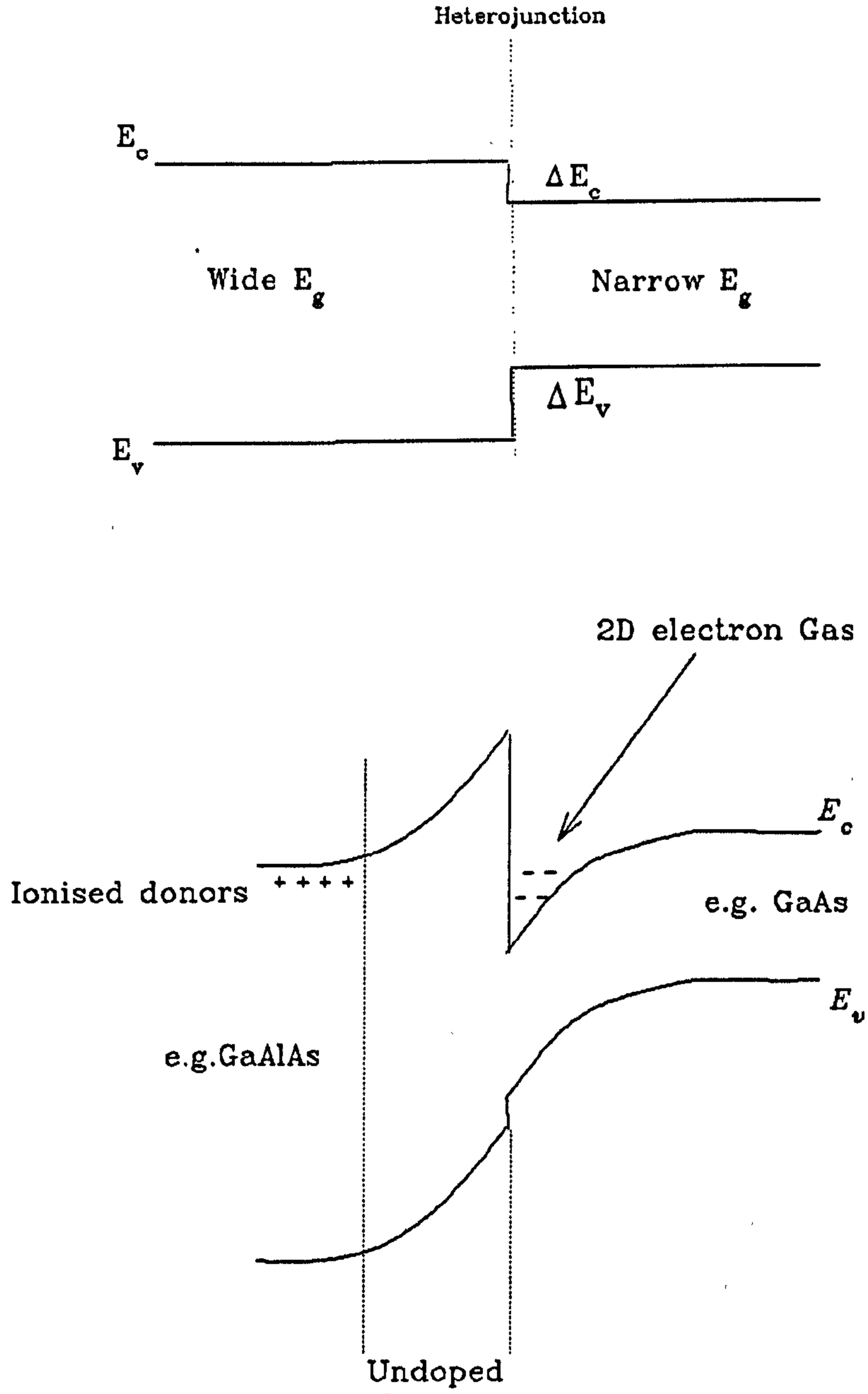


Fig 5.3 Band diagram for a remote doped 2DEG formed in a GaAs channel. The upper diagram represents an undoped heterojunction.

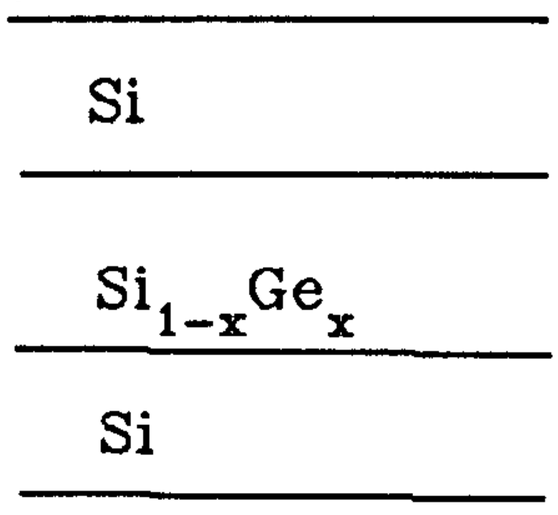
doped spacer region. At zero bias, the Fermi level must be constant throughout the structure, which requires charge transfer. This occurs by depletion at the n^+ (p^+) junction, as expected for a pn junction, and by electrostatic attraction between carriers at the interface and the depleted parent atoms in the doped region. Thus, the minimum requirements for successful remote doping are a heterojunction, an abrupt doping transition and a suitable band offset, $\Delta E_{c,v}$, at the heterojunction interface, greater than a few kT to allow for finite quantum well width.

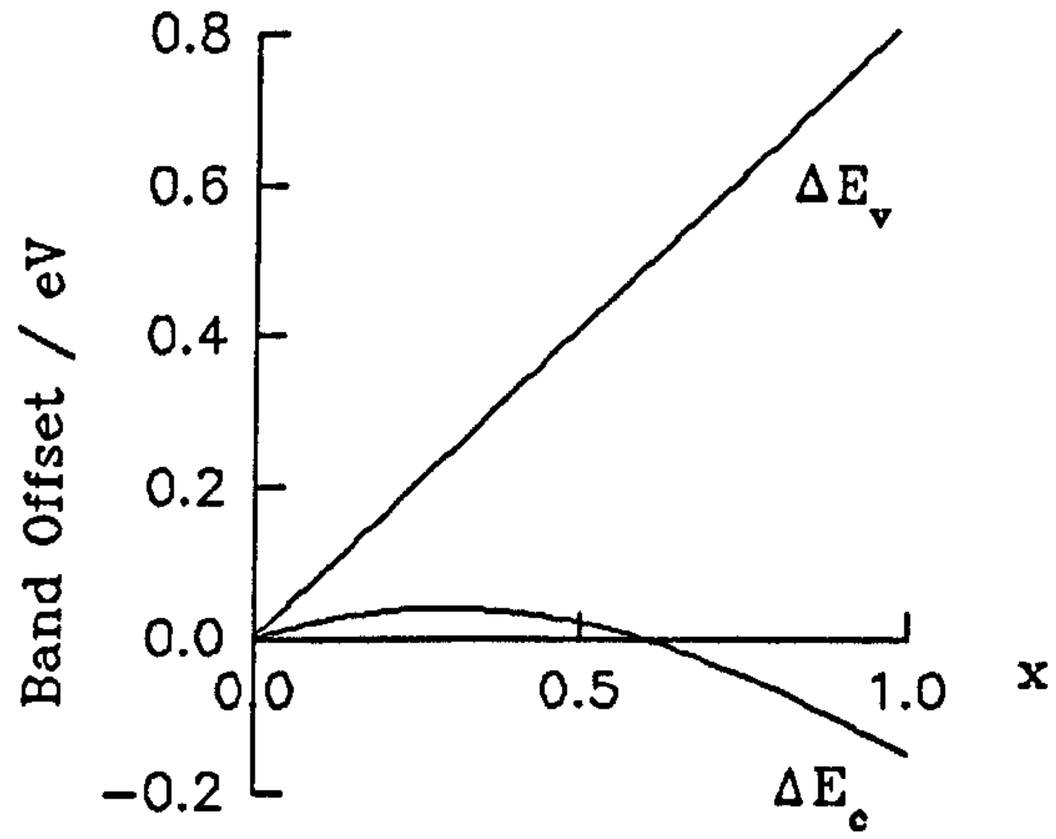
5.1.3 Remote Doping In SiGe

Initial remote doping experiments using a buried $Si_{0.8}Ge_{0.2}$ layer, strained to <100> Si, by People et al 1984, produced a 2D hole gas but no evidence of an electron gas. This suggested that the band offset is accommodated predominately at the valance band. Self-consistent pseudopotential calculations for this system by Van de Walle and Martin 1986 and People and Bean 1986 predicted very small ΔE_c, but ΔE_v of 170meV and 150meV respectively. These are in reasonable agreement with experimental measurements by Ni and Hansson, 1990, using X-ray photoelectron spectroscopy, and Khorran et al 1991, as deduced from the temperature dependence of heterojunction I-V results.

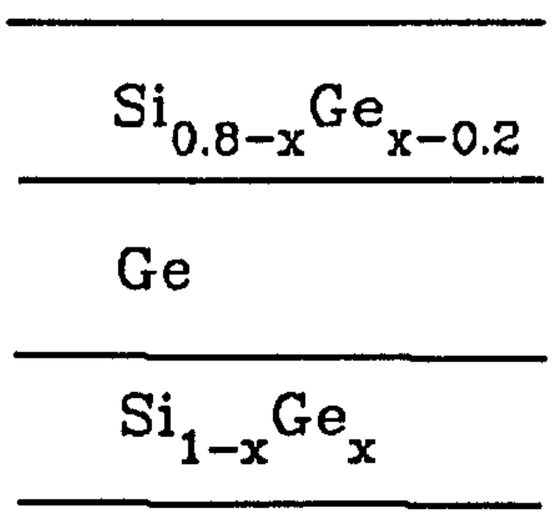
Fig 5.4 shows the band offsets for the possible SiGe based systems generated from the equations of Van der Walle and Martin, 1986. Clearly, it is possible to create a remote doped electron structure using SiGe technology. A Si/Si_xGe_{1-x} ($x > 0.5$) system strained to match a relaxed SiGe $x = 0.3$ buffer has suitable ΔE_c for a Si channel to 2DEG. The first experimental achievements of this came from Abstreiter et al 1985. The difficulty with this structure is the very high threading dislocation density ($\sim 10^{12}$ cm^{-2}) in the channel propagating up from the relaxed buffer layer. A number of publications over the next six years reported minor growth modifications resulting in a

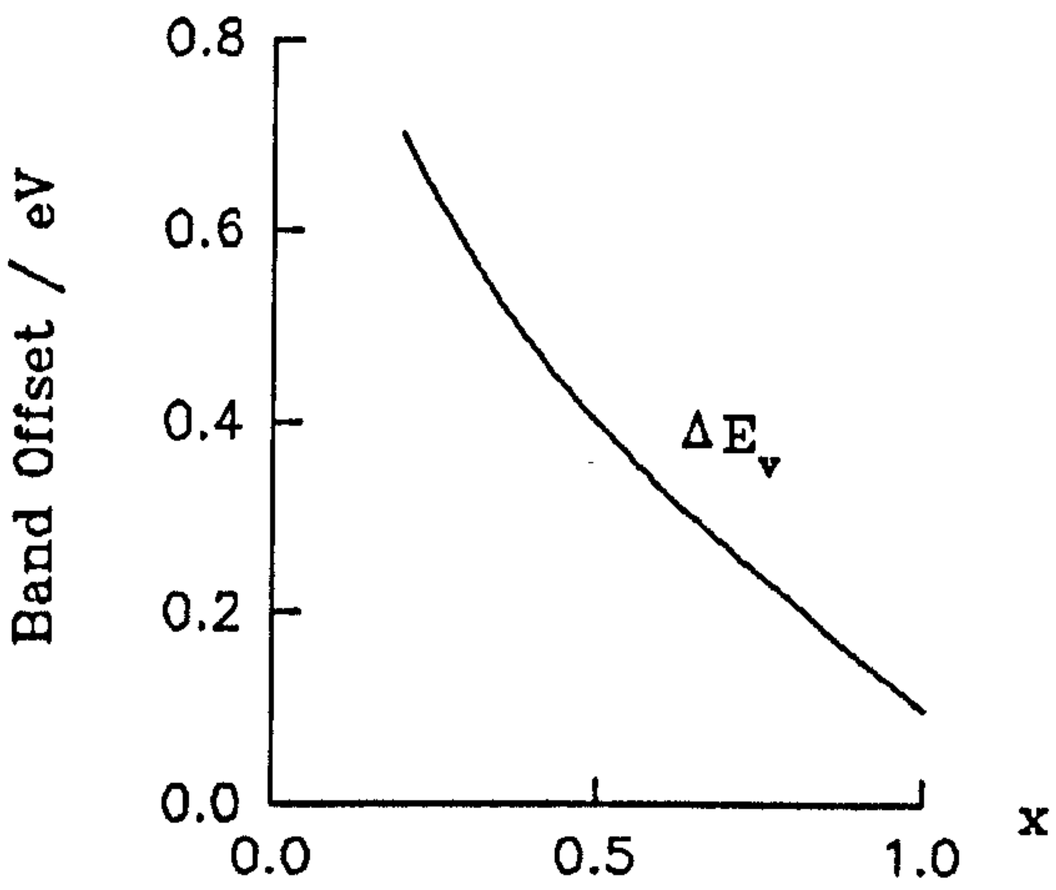
(a)



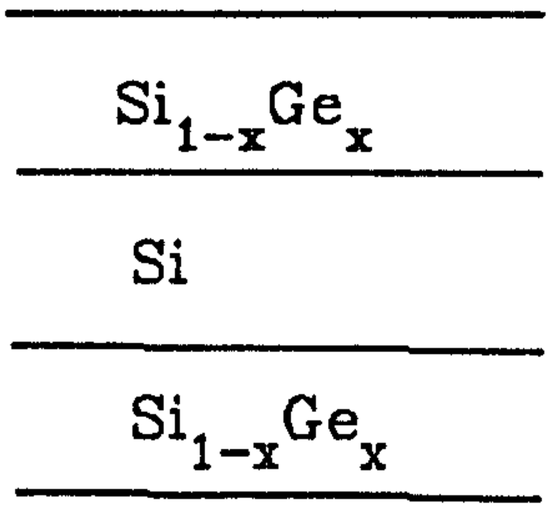


(b)





(c)



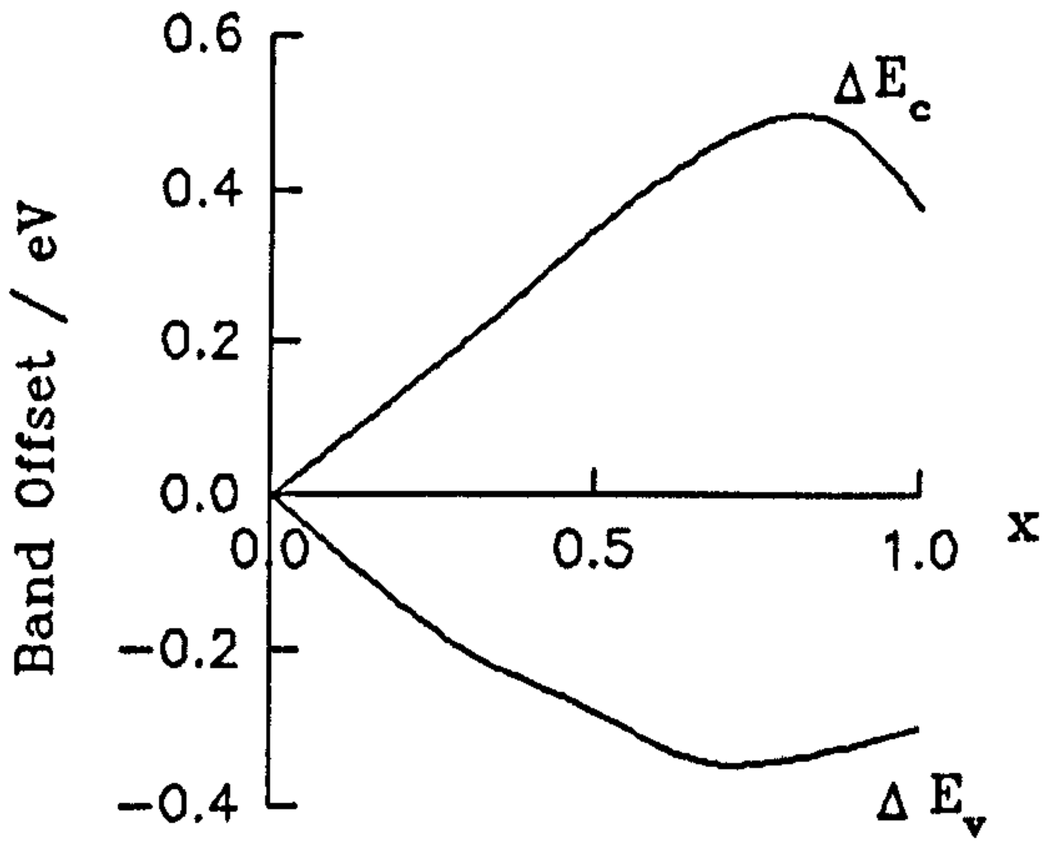


Fig. 5.4 Approximate band offsets for three SiGe heterojunction systems. (a) the SiGe 2DHG, (b) the Ge channel 2DHG and (c) the Si channel 2DEG.

4.2K mobility, μ, of 1.7 x 10^4 cm^2 $V^{-1}s^{-1}$ (Schuberth et al, 1991). The problem was solved by Mii et al, 1991, who used a graded Ge concentration buffer (graded at 10 % per um) and achieved $\mu(4.2K) = 9.6x10^4$ cm^2 $V^{-1}s^{-1}$. An alternative low dislocation SiGe buffer had been reported by Powell et al, 1990, who grew step graded layers of increasing Ge content, each being grown to near the metastable t_c, then in-situ annealed to cause relaxation. Thus, the lattice mismatch of the subsequent deposition is relieved. The misfit dislocation density is low because the high temperature anneal (900^0C) produces long misfit dislocations (of order millimetres) to relieve strain. The principle of the slowly graded buffer is less clear. Subsequently, Schaffler et al 1992 used the graded buffer technique to achieve $\mu(4.2K) = 1.7 \times 10^5$ cm^2 V^{-1} s^{-1} and to provide strong evidence that the mobility, at least below this value, was limited by dislocations. This material allowed the first observation of the FQHE in a Si based system.

SiGe on Si 2DHG structures have been reported by number of groups using various growth techniques. A summary of the results is given in Table 5.1. The most interesting features of these results are the low values of $\mu(4.2K)$ and their consistency. In the absence of calculations for the scattering rates in this system, Mishima et al 1991, suggested that mobilities were limited by background ionised impurity scattering from B in the channel, whereas Nutzel et al 1992, and Venkataraman et al 1993 suggest that the limitation was alloy scattering in the channel, which would explain the consistently low values. As will be shown in Section 5.4.7, a threefold improvement in $\mu(4.2K)$ compared to these values is possible.

5.2 THEORY

The electrical effect of impurities and crystalline defects on low temperature 2D parallel transport is dissimilar to the recombination/generation in 3D devices as

μ(4K) / $cm^2V^{-1}s^{-1}$	Carrier concentration / cm^{-2}	Ge Fraction	Structure Type	Growth Technique	Reference
3000	$7x10^{11}$	0.2	Double Symmetric	MBE	People et al, 1985
≈4200	$8x10^{11}$	0.12	Double Symmetric	UHV CVD	Wang et al 1989
4000	$1x10^{12}$	0.2	Normal	MBE	Mishima et al 1990
2500	$5x10^{11}$	0.15	Normal	Rapid Thermal CVD	Venkataraman et al, 1991
3400	$3x10^{11}$	0.28	Normal	MBE	Nutzel et al 1992

Table 5.1 Published electrical transport measurements on $Si_{1-x}Ge_x$ channel, remote doped 2D hole gases, indicating the highest mobilities obtained.

discussed in Chapter Four. Therefore, in order to discuss fully the transport measurements of this chapter and their interpretation, it is necessary to introduce the concept of electron(hole) localisation and to comment on 2D carrier scattering mechanisms. Only brief qualitative descriptions will be given.

5.2.1 Localisation

The theory of localised wave functions in solids was given initially for highly disordered amorphous systems by Anderson, 1958 and developed by Mott, 1974. Short range disorder can be viewed as a perturbation on the Hamiltonian for a system leading to band broadening, whereby energy states "tail" into the bandgap. This leads to an increased probability of a carrier being localised at a disorder site. This localisation is termed strong or weak, both of which influence transport processes at low temperatures.

5.2.1.1 Strong Localisation

Anderson defined a strongly localised state as one with zero diffusion at T = 0K such that at time $t\rightarrow\infty$, the electron has a finite probability, proportional to $e^{-2\alpha r}$, of remaining within a distance r. For extended states, the DC conductivity is finite at T = 0, whereas at localised states it is zero. Note that if sample dimension, L, is less than the localisation length, $1/\alpha$, then the concept of localisation is not useful.

Wavefunction localisation due to vertical disorder is illustrated in Fig. 5.5. Disorder gives rise to an energy E_C in the density of states below which all states in the band are localised and above which all are extended. E_C is termed the mobility edge, since the position of E_F relative to E_C will determine the T=0 conductivity, i.e. 0 or finite.

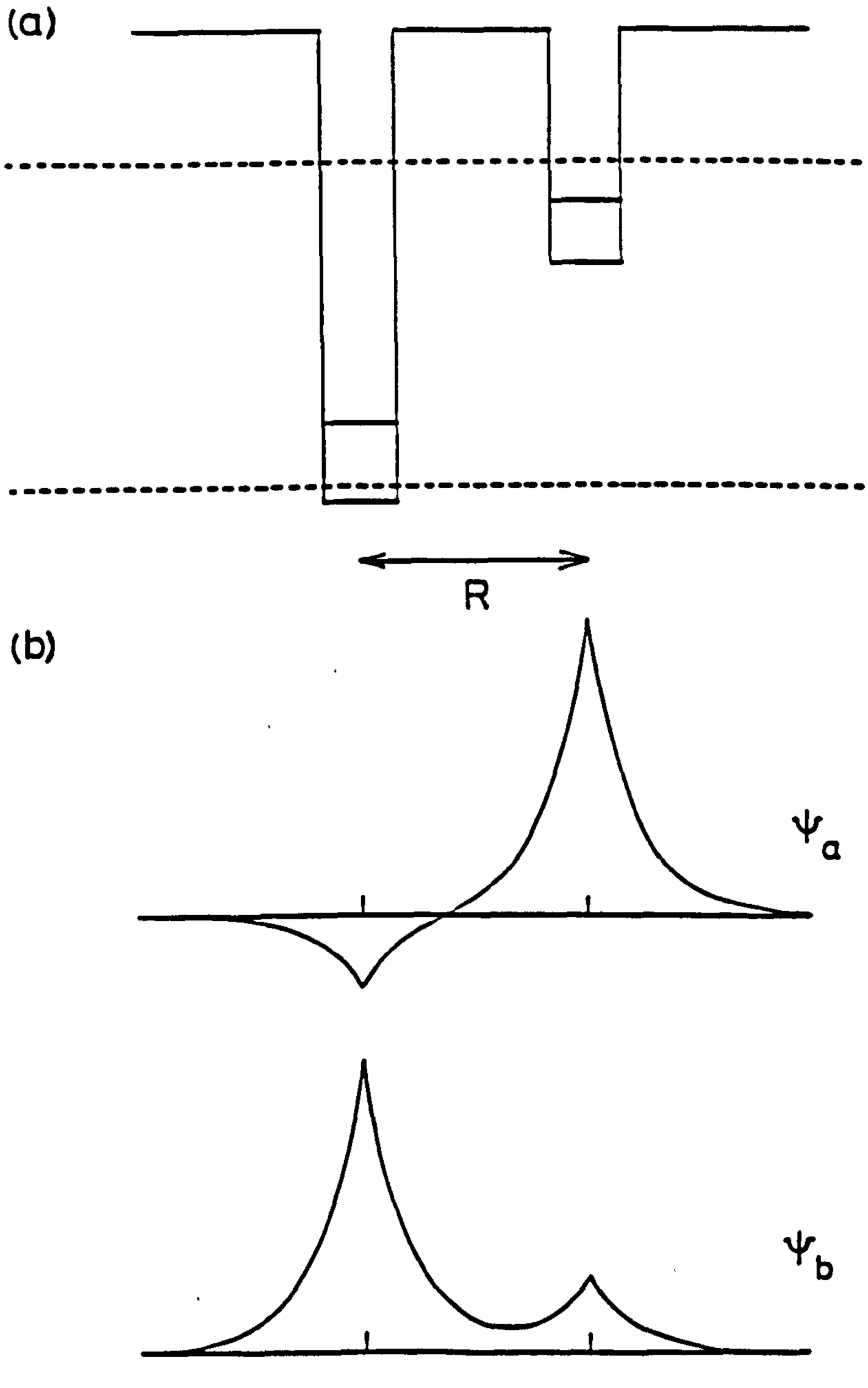


Fig 5.5 (a) Schematic of perturbed potential wells , due to vertical disorder.
(b) Symmetric and anti symmetric wavefunctions, localised by this disorder
(After Whall, 1993)

In the strongly (exponentially) localised systems, conduction above T = 0 is phonon assisted: this conduction is either variable range hopping between sites at or near E_F, or is activated hopping between nearest neighbours. The hopping conductivity is proportional to the hopping probability, P, which is governed by the distance between sites, R, the inverse Bohr radius, α, and the mean energy required for a site transition. It is given generally by

$$\sigma \propto P \propto e^{\left(-2\alpha R - \frac{\Delta E}{kT}\right)}, \qquad 5.1$$

and the optimum value of R is found at dP/dR = 0. In 2D we find that

$$\sigma \propto P \propto e^{T^{-1/3}}. \qquad 5.2$$

This result can be used to distinguish 2D and 3D systems (where $\ln \sigma \propto T^{-1/4}$), or indeed to establish the existence of a Coulomb gap due to electron-electron interactions (where $\ln \sigma \propto T^{-1/2}$). At higher temperatures carriers are thermally activated above E_C and this activated conduction follows an $e^{-\left(T_0/T\right)}$ relation.

In practice, the hopping conduction regime only occurs for very low temperatures in crystal semiconductors and the data available in a ρ versus T plot is often not strictly adequate to justify a given coefficient. By assuming a relation $\rho \propto \exp\left(T_0/T\right)^n$ (where T_0 is a transition temperature for activated conduction) and defining a function $G(\rho, T)$ as $\partial \ln \rho / \partial(1/T)$, it can be seen that a plot of log G versus log T yields gradient (1-n). In practice, G is found by fitting a polynomial to the data ($\ln\rho$ versus 1/T), thence manual differentiation as performed in the Zerbst analysis. This will be used later.

5.2.1.2 Weak Localisation

Weak localisation occurs when the probability of an electron remaining at a disorder site follows a power law, rather than exponential. This evinces itself in otherwise metallic systems (i.e. $\frac{d\sigma}{dT} < 0$) at very low temperatures when $\frac{d\sigma}{dT} > 0$.

Two important results for 2D systems can be readily derived using scaling theory, Abrahams et al 1979. The first is that all states in 2D (or 1D) are localised, i.e. there can be no 2D metal. The second is that a logarithmic correction term to the Boltzmann equation is found necessary for 2D systems. A physical description of weak localisation deriving from single electron self-interference has been provided by Bergmann, 1984. In addition, a further logarithmic correction has been shown to arise from disorder enhanced electron-electron interactions by Al'tschuler et al, 1980.

Expressions for localisation corrections to the 2D conductance and magnetoresistance have been summarised by Lee and Ramakrishran 1985, and a description of quantum interference effects given by Al'tshuler and Aranov 1985. A useful property of readily obtainable experimental data made clear in these analyses is that, whilst electron-electron interactions (via $N(E_F)$) and weak localisation (via D), both contribute to the conductivity, the Hall coefficient, R_H, is sensitive only to interactions. The two effects may, therefore, be quantitatively distinguished by experiment.

These theories have been successfully applied by the author and co-workers to SiGe 2DHG samples with 4K mobilities $\approx 2 \times 10^3$ cm^2 V^{-1}s^{-1} (see Section 5.4.5). This particular work, however, does not form a primary part of this study, so the details of the theory will not be fully developed here (instead, the reader is referred to Emeleus et al 1992 and 1993). It does, however, provide significant evidence regarding material quality. As such, it is referred to in Section 5.4 regarding the effects of, and possible

sources of, localisation in the 2DHG. I would merely note here, that the theories were constructed in the limit of weak disorder i.e. $k_F l >> 1$ but are successful in describing these low mobility samples with $k_F l \approx 1$. This may prove to have significance for future theoretical developments.

5.2.2 Scattering

Electrons in a lattice at room temperature exchange energy with the lattice in order to maintain thermal equilibrium via inelastic phonon scattering. (If the E field applied accelerates electrons such that this mechanism cannot maintain thermal equilibrium, then the electrons are termed "hot", with energy characterised by an electron temperature.) Remote doping techniques will not increase 2D mobilities by orders of magnitude over equivalent bulk values at room temperature since phonon interactions are dominant. At liquid He temperatures, however, phonon scattering is negligible compared with elastic scattering via lattice imperfections, except in very high mobility samples, Foxon et al 1989.

In this work we are concerned with the growth requirements for high mobilities at $T < 50K$. As such, the scattering mechanisms of interest are via impurity atoms, disorder due to random fluctations in alloy composition and perturbations (roughness) at the heterojunction interfaces. Much of the framework for treatments of 2D elastic scattering was developed for Si inversion layers and is reviewed in Ando et al 1982 and subsequently was developed for III-V heterojunctions by various authors (see, for instance, Harris et al 1989).

The details of the calculations for each scattering mechanism are not of interest here and are fully described, for the structures of this Chapter, by Emeleus et al 1993, from which Fig. 5.6 is taken. However, it is necessary understand the effect of 2D

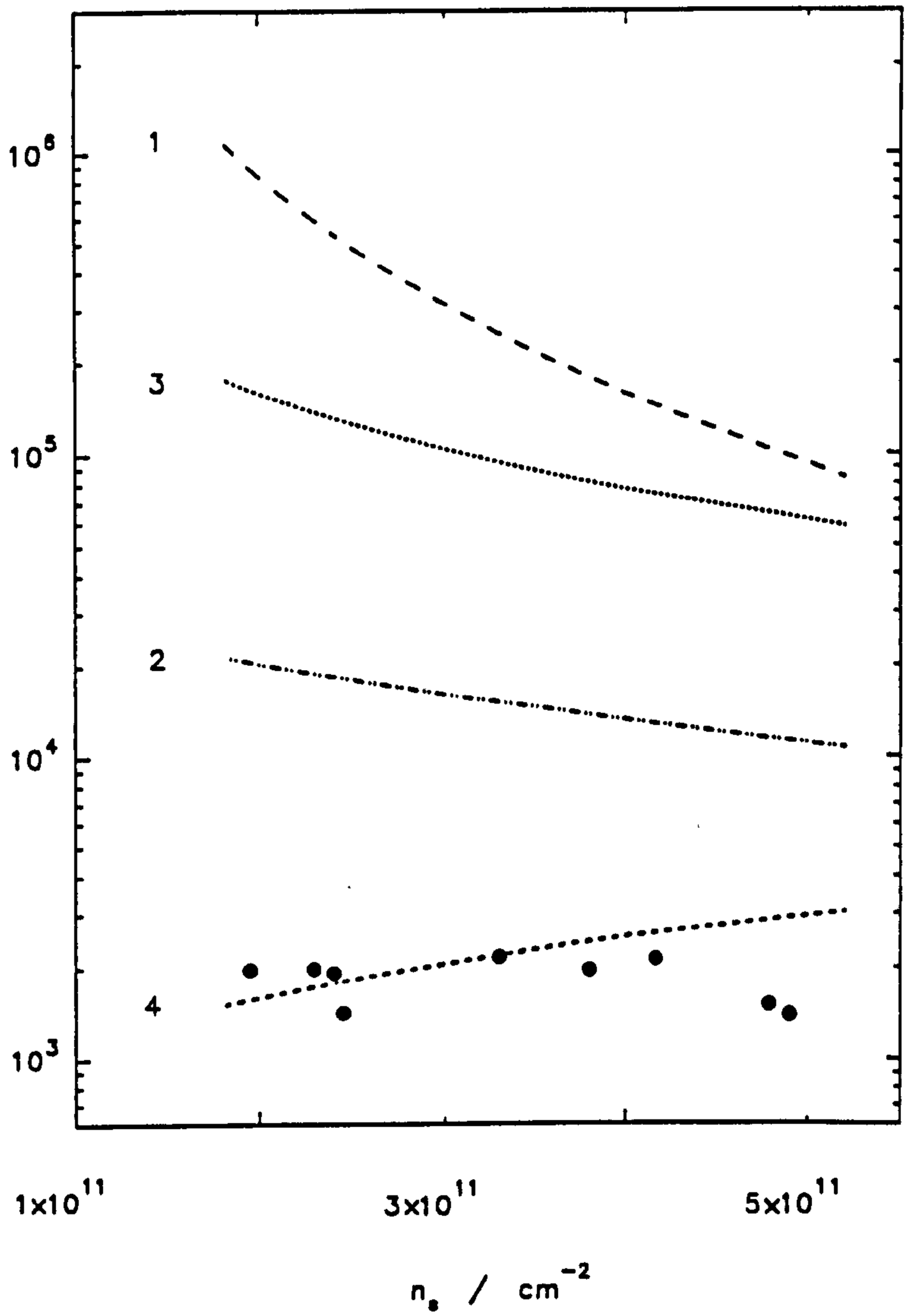


Fig. 5.6 Variation of 2DHG mobility with sheet concentrationfor the following scattering mechanisms: (1) Remote impurities, (2) Interface roughness, (3) Alloy scattering, (4) Interface impurities (ni=$2x10^{11}$ cm^{-2}) (After Emeleus et al 1992)

carrier concentration on the scattering mechanisms relevant to this study. Fortunately, this can be achieved qualitatively by considering just one equation, the relaxation time, τ, approximation, Butcher 1986:

$$\frac{1}{\tau} = \frac{A}{(2\pi)^2} \int P(kk')(1 - \cos\theta) d^2k' \qquad \textbf{5.3}$$

where A is the area of the 2D gas, P(kk') is the probability of scattering from k to k' (given by Fermi's Golden Rule) and the 1-cosθ term implies that only momentum changes in the plane of current flow are important. Note that the conductivity, σ, for a degenerate gas is given by $\sigma = \frac{ne^2\tau(E_F)}{m^*}$. In a remote doped 2D system $\theta \rightarrow 0$, so 1-cos$\theta \rightarrow 0$, so $\tau \rightarrow \infty$. This is why remote doping is effective.

Interface Roughness

Well thickness variations lead to a broadening of the sub bands, i.e. more possible kk' transitions, therefore more scattering, so τ decreases with increasing roughness. Higher n_s reduce the well width (for a triangular well), so increasing the sub band broadening for a given well thickness variation.

Ionised Impurity Scattering

The relaxation time is inversely proportional to the density of ionised impurities, so this mechanism can be strong, but is ideally weak due to the remote doping effect. In a degenerate system, however, the conductivity is governed by $\tau(E_F)$, i.e. the carrier velocities are high and P(kk') is small (c.f. Rutherford scattering). Thus, for high n_s, scattering is reduced.

Alloy Scattering

Unscreened alloy scattering is independent of n_s because the perturbing potential is very short range.

The effect of screening, whereby carriers redistribute to minimise energy, is neglected in the forgoing discussion, but has been accounted for in the calculations represented in Fig. 5.6 and introduces an n_s dependence for alloy scattering.

Neutral Impurity Scattering

A scattering mechanism little discussed in the literature, neutral impurity scattering (NIS), was thought at an early stage of this work to be dominant in this system. It was later found unnecessary to invoke NIS to explain experimental observations. However, the growth studies resulting from this line of inquiry directly led to 2DHG mobilities significantly higher than previously reported, so it warrants a brief description. NIS has been considered in 3D (by Erginsoy, 1950), as elastic scattering from a neutral H atom: these calculations predict a temperature independent mobility. In 1956, Sclar pointed out that, at low T, the H atom can ionise by acquiring a weakly bound electron, the D^- state, or a hole, the A^+ state, thereby exerting a short range attraction on free carriers. This would lead to a transition between ionised and neutral scattering and hence a resonance condition, such that a minimum would exist in the μ-T relationship, of the form

$$\mu \propto A\sqrt{T} + \frac{B}{\sqrt{T}} \qquad 5.4$$

McGill and Baron 1975 returned to this problem using acurately determined phase shifts for electron scattering from neutral H. This analysis also predicted a minimum in μ-T at

$T = \frac{km_h\kappa^2}{10E_0m^*}$, (where E_0 is the ground state binding energy, k is the Boltzmann constant and κ the dielectric constant), i.e in the range 1 to 25K depending on E_0. Such a minimum was observed in P doped Si by Norton et al 1973, who concluded that it resulted from resonant NIS. A review of H-like impurity centres in semiconductors is given by Gershenzon et al 1985. Analytical expressions for 2D NIS have not been

developed to our knowledge, but, qualitatively, a similar resonant condition would seem possible.

5.3 EXPERIMENTAL DETAILS

5.3.1 The Hall Effect

Classically, the Hall effect refers to the effect of the Lorentz force induced by application of magnetic field, B_z, perpendicular to the current I_x. The Hall field balances this force i.e

$$E_H = \frac{V_H}{t} \quad \text{and} \quad \frac{E_H}{J_x B_Z} = \frac{1}{nq} \quad \text{for uncompensated material.} \qquad \textbf{5.5}$$

We define 1/nq as R_H, the Hall coefficient. Since $\rho = \frac{1}{qn\mu}$ then $\mu = \frac{R_H}{\rho}$ **5.6**

Thus, by measuring $\rho(B=0)$ and V_H, then μ and n may be found. The Hall mobility is distinguished from the drift mobility by a scattering constant, r, contained within the definition of R_H:

$$R_H = \frac{r}{q}\frac{p - b^2 n}{(p + bn)^2} \quad \text{, where } b = \frac{\mu_e}{\mu_h}. \qquad \textbf{5.7}$$

This constant, r, accounts for the scattering mechanism via the mean free time between scattering events, $r = \frac{<\tau^2>}{<\tau>^2}$ and is of the order 1. Whilst r may be calculated for a known mechanism, to apply this value to an experimental result requires an a priori knowledge of the conduction processes, which is often the aim of the experiment. It is particularly difficult to choose r when two or more scattering mechanisms are present and have differing temperature dependencies. (Note that this can lead, in the worst

case, to 100% difference between μ (Hall) and μ (drift).) However, for degenerate systems, the relaxation time is single valued, so r=1.

For Hall bar studies, the tensor nature of ρ must be considered (and it is highly useful so to do), but it is not necessary for the results of this thesis, in which Hall cross results are quoted. A common complication in epitaxial Hall measurements is the existence of parallel conduction paths. The two layer problem has been treated by Larrabee and Thurber 1980 (and many other authors) yielding

$$\mu_H = \frac{\mu_1^2 n_1 + \mu_2^2 n_2}{\mu_1 n_1 + \mu_2 n_2} , \qquad 5.8$$

a solution of which requires knowledge of one of the paths from independent measurements. For uniformly doped epilayers, high resitivity substrates of opposite polarity are used to provide pn isolation. The use of controlled depth contacts is useful in this case i.e implantation. For these remote doped samples, p^- substrates were used and possibly contacted, but conduction for T > 50K is dominated by the doped region for carrier transfer to the quantum well; indeed, the μ-T curves at high T (> 70K), are fully explicable in terms of the 3D behaviour of this p^+ layer, which is inevitably contacted. Delta layers have been used by other workers as the charge source, which deplete fully at 300K. However, this was felt to represent an unnecessary complexity and added constraint upon the growth schedule at the development stage. This approach was later proved to be justified, but does mean that no useful information was obtained above 50K i.e until full 3D carrier freeze out. That no other conduction paths were present for T < 50K has been confirmed in these samples by the observation of single period Shubnikov de Haas oscillations (Emeleus et al, 1993).

The sample configuration used was the etched van der Pauw cross (as described in Chapter Three) with current passed through diametrically opposite contacts and V_H measured at the remaining two. A number of asymmetric voltages are present in these measurements, as detailed by Wieder, 1979. These are removed by interchanging IV

measurements and by B field reversals. The exception is the Ettinghausen voltage, inherent to electron flow in a magnetic field, but this are always much less than V_H. The extraction of sample resistivity and Hall coefficient from the various configuration measurements is well-known and is given in Van der Pauw, 1958. The symmetry of sample is characterised by the "f" factor, ideally 1, which was calculated in each sample configuration at each measurement, and any results with $f < 0.9$ considered invalid.

The assumption was made, that for $B < 0.5$ T, these are low field measurements, i.e $\mu B << 1$, and that the magnetoresistance is zero. Any discrepancy between ρ (B=O) and ρ for B non zero is ignored. From Emeleus et al, 1992, this assumption can be seen to introduce only approximately 5% error for the fields used.

5.3.2 Cryogenics

Three cryostats were used at various times in these studies - a closed cycle cryostat (CCC) with T_{min} ~ 13K, a He_4 bath cryostat (BC) with T_{min} ~ 5K and a continuous flow cryostat (CFC) with T_{min} ~ 3.4K.

The CCC is an Air Products Displex with a RhFe thermocouple as temperature sensor and a 0.4T permanent magnet. The principle of operation is that of a domestic refrigerator, but using He as the coolant. The BC is a modified Thor cryogenics system with a 0-0.5T electromagnet and a capacitance temperature sensor, a Lakeshore capacitance controller for heater control and a RhFe resistor as sample temperature sensor. The measuring electronics is common to these two systems and is described in detail by Biswas, 1992. The CFC is an Oxford instrument cyrostat with a Si diode in constant current mode as temperature sensor (diode supplied and calibrated by Institute

of Cryogenics, Southampton University) and RhFe thermocouple as sensor for an Oxford ITC4 temperature controller. The measuring electronics system was constructed by the author and is depicted in Fig 5.7. The measurements were controlled and data manipulated by a BBC micro computer through an IEEE interface, using software written by the author. The data was transferred to a PC compatible format and analysed using a Jandell software package Sigma Plot.

5.4 RESULTS AND DISCUSSIONS

5.4.1 Sample Design

This study relates to the formation of 2DHGs in SiGe strained to Si. Such structures may be of three types - normal, inverted and symmetrical, i.e. with a 2DHG formed at the upper, lower or both heterojunctions respectively. The considerations for structure design are the carrier population in the well (n_s), Ge concentration, B doping concentration, spacer widths, channel width and choice of interface. The carrier population and sub band occupancy can be calculated in terms of these other parameters. The well shape and carrier concentrations, n_s, are mutually dependent functions, making a simultaneous solution of Poisson's and Schrodinger's equations by numerical methods a useful approach. There is, however, an approximate method for obtaining an estimate of n_s which was used by the author to design the structures used in this work.

The numerical technique applied in this study was proposed for GaAlAs HEMT design by Delagebeadeuf and Linh 1982. The sheet concentration in the quantum well, n_s, is found by a numerical comparison between the density of states in the well and the depletion approximation as applied to the doped Si region.

a) Depletion Approximation

A formulation for n_s can be found from the depletion approximation applied to the doped Si region. Defining the system from Fig 5.8, taking Poisson's equation for

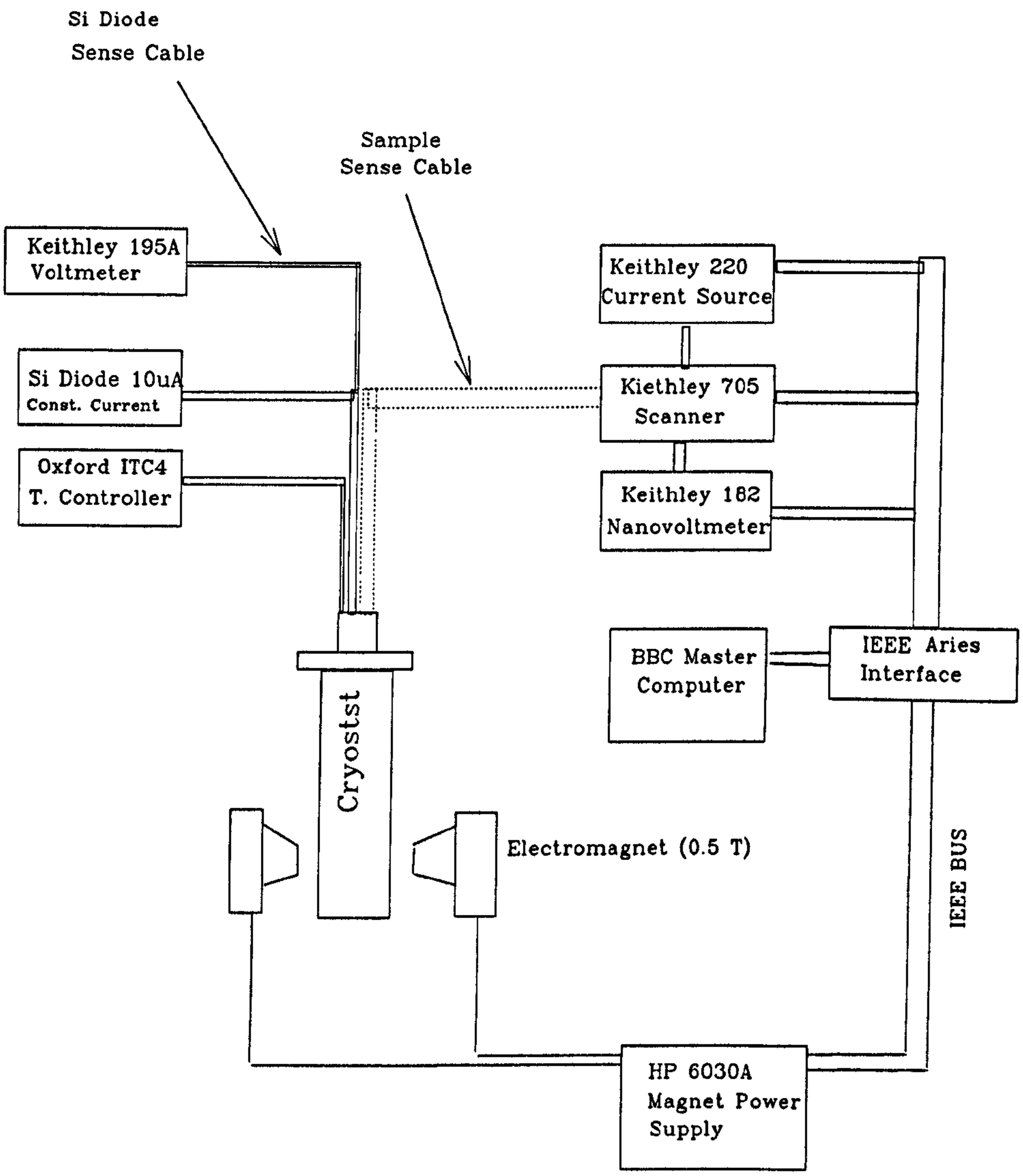


Fig. 5.7 Schematic of measuring system for continuous flow cryostat.

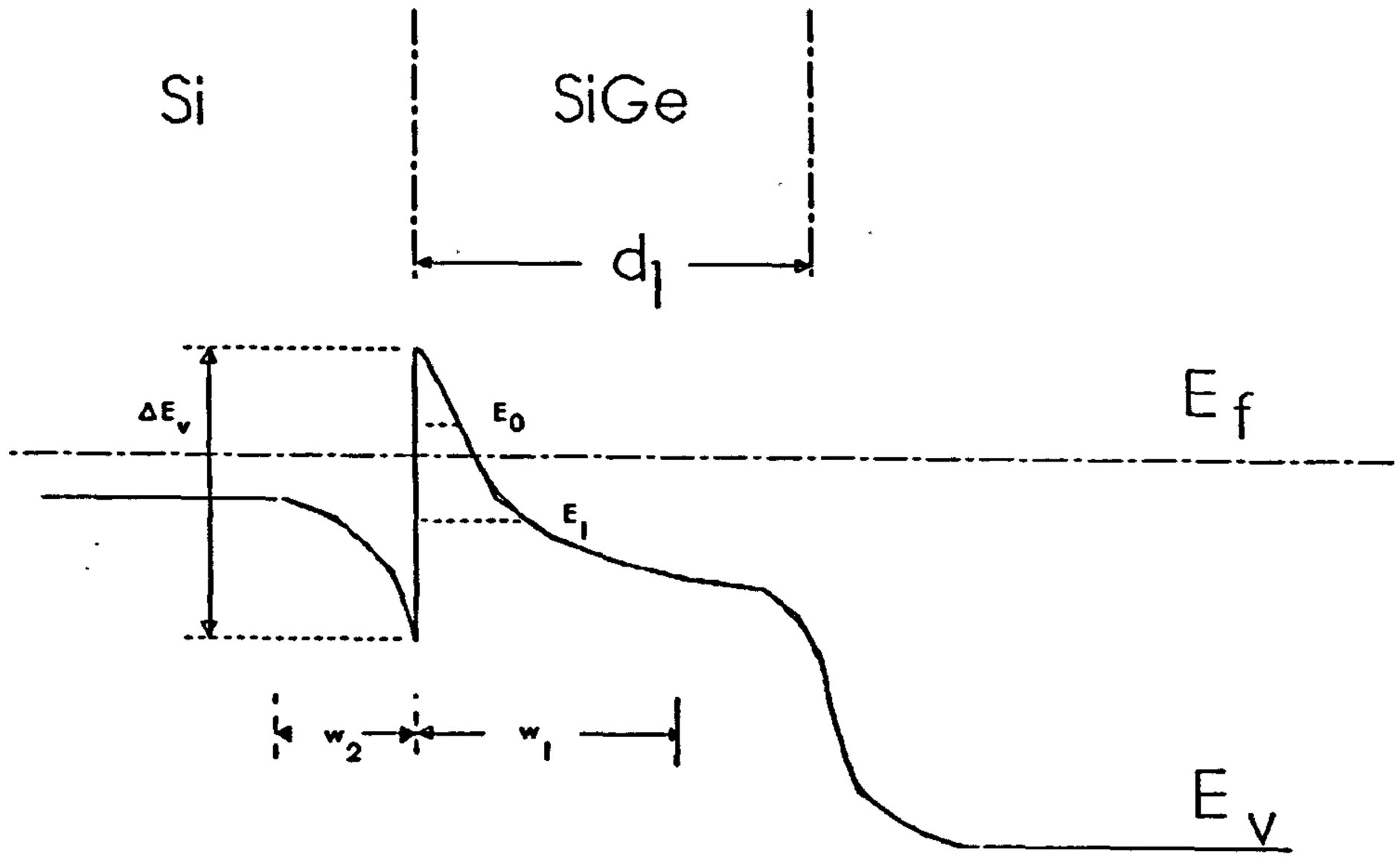


Fig. 5.8 Schematic of the 'normal' SiGe quantum well. Parameters for the sheet concentration modelling are defined in this figure.

the B-doped Si region (defining x=0 at the Si/SiGe interface), then the field at the interface, F_i, is

$$F_i = \frac{-q}{\varepsilon_2}\int_0^{-w_2} N_2(x).dx \qquad \textbf{5.9}$$

In the spacer layer, thickness t, $N_2(x) = 0$, and for $x < -t$, (i.e. the B doped region) $N_2(x)=N_2$, therefore

$$F_i = \frac{-qN_2}{\varepsilon_2}(t - w_2) \qquad \textbf{5.10}$$

In order to account for band bending, ν, in the Si, we define $V(0) = 0$. Thus $\nu = V(-w_2)$, which is found from Poisson's equation using the result for F_i found above, viz

$$\nu = \frac{qN_2}{2\varepsilon_2}\left(w^2 - t^2\right) \qquad \textbf{5.11}$$

Rearranging for w_2 and substituting in equation **5.19** yields

$$\varepsilon_2 F_i = -qN_2 t + \sqrt{2qN_2\varepsilon_2\nu + q^2N_2^2t^2}\,, \qquad \text{and, by Gauss' Law, } \varepsilon_2 F_i = qn_s. \qquad \textbf{5.12}$$

However, ν depends on the Fermi energy, E_F, i.e. $\nu=\Delta E_c - E_F$. ΔEc can be taken from Van de Walle and Martin, but E_F is an unknown. Therefore, an alternative formulation for n_s is required in order to find E_F and can be found from the 2D density of states in the SiGe well.

b) 2D DOS In The SiGe Well

Assuming a constant electric field, F, in the well (hence a triangular well), which yields

$$E_n = \left(\frac{\hbar^2}{2m_l^*}\right)^{\frac{1}{3}}\left(\frac{3}{2}\pi qF\right)^{\frac{2}{3}}\left(n+\frac{3}{4}\right)^{\frac{2}{3}} \qquad \text{for n=0,1} \qquad \textbf{5.13}$$

where only F_i, the field at the interface, is an unknown. Poissons equation for the SiGe channel can be written as

$$\frac{dF}{dx} = \frac{-q}{\varepsilon}(n(x) - N_d) \qquad \textbf{5.14}$$

For the majority of structures in this study d_1 is 50nm, which justifies taking $F(d_1) = 0$. Thus, integrating yields

$$F_i = \frac{q}{\varepsilon_1}(n_s - qN_d d_1) \approx \frac{q}{\varepsilon_1} n_s \qquad \text{so} \qquad E_n \propto (n_s)^{\frac{2}{3}} \qquad \textbf{5.15}$$

The 2D DOS for one sub-band is given by $D = \frac{qm^*}{\pi\hbar^2}$. So, using Fermi Dirac statistics, an expression for n_s is found:

$$n_s = D\int_{E_1}^{E_0} \frac{dE}{1+e^{\frac{q(E-E_F)}{kT}}} = D\frac{kT}{q}\ln\left[\left(1+e^{q/kT(E_F-E_0)}\right)\left(1+e^{q/kT(E_F-E_1)}\right)\right] \qquad \textbf{5.16}$$

In order to use this equation, both n_s (hence E_0 and E_1) and E_F are required.

Thus, n_s can be found from the depletion approximation for arbitrary E_F and can then be used to find E_n for equation **5.16**. This can then be used to recalculate n_s from the 2D DOS to yield a new value for n_s. Using a BASIC programme written by the author, this process is repeated for incremental steps of E_F, 10meV, until the values of n_s agree to within 10%. E_F is then reset by subtracting 10meV, and iterating increments of 1meV, until n_s agrees to 1%. The numerical solutions were compared with published results from MBE and CVD growth (Fig 5.9). The agreement is adequate for structure design and this was a useful guide throughout this work.

Note that these calculations indicate ground state occupancy only for the structures studied - which has been confirmed by the observation of single frequency Shubnikov de Haas oscillations for the structures used in this study. The scatter in the data of Fig. 5.9 is surprising. Growth parameters e.g. growth rates, doping concentrations and alloy concentration for low dimensional structures are, normally, accurately characterised and further, the model is not critically sensitive to these parameters. The assumption of a triangular well leads to an overestimate of sub-band energy and, thus, an underestimate of n_s. This, however, is not the sole reason for the general underestimation of Fig. 5.9. The discussion in Section 5.4.5 suggests an additional explanation for this underestimation and scatter, from which a self consistent model incorporating experimentally observed features was developed, largely by M J Kearney and C J Emeleus.

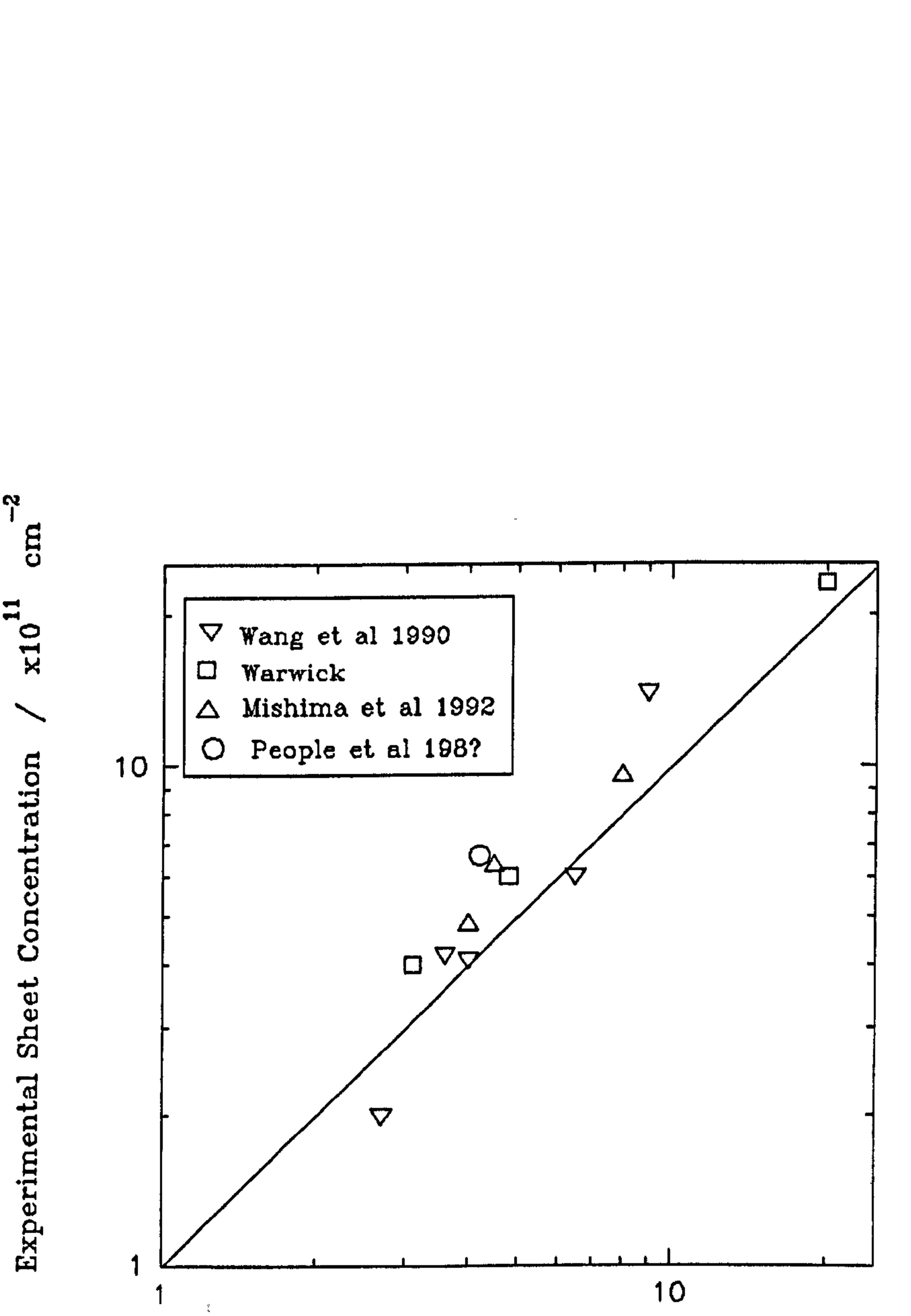


Fig. 5.9 Comparison between theoretical and experimental carrier concentrations. The straight line represents agreement between model and experiment.

5.4.2 Initial Studies

The progress of the growth studies, with regard to the 4K mobilities achieved, is summarised in Fig. 5.10. This figure is provided as a guide, for the reader, to the remainder of these results. With reference to this figure, there are four distinct phases to these studies that have increased the 4K mobility from ~100 $cm^2V^{-1}s^{-1}$ to ~9,000 $cm^2V^{-1}s^{-1}$, i.e. a greater than twofold increase over the highest mobility previously reported. Each of the improvements in mobility were the result of changes in the growth conditions. The remainder of this chapter describes these four phases and provides some possible explanations for the beneficial effects.

In the early stages of this work (Phase A), approximately thirty B doped SiGe on Si remote doped structures were grown in various configurations - two channel symmetrically doped, multi-quantum well, and single quantum well. Common to all of these were mobilities ~100 $cm^2V^{-1}s^{-1}$ and high resistivities for T < 10K. Two processes were, initially, considered as causes for these effects: the diffusion of B into the well (ionised impurity scattering) or highly disordered interfaces (interface roughness scattering). However, neither explanation appeared probable given the previous work by the group on B doping and SiGe/Si interface characterisation, so these results were highly puzzling. Following the work of Mishima et al 1992, efforts concentrated on the single channel normal quantum well with Ge fraction nominally 0.2, as shown in Fig 5.11. B segregation from the surface adlayer could be expected to present difficulties for the inverted structures where the well is formed at the lower interface. Inverted structures were grown intermittently throughout this study and, in all cases, were found to have mobilities significantly less than for normal structures. This has not been investigated in detail.

Assuming, at this stage, that atomic scale roughness was limiting the mobilities, growth interruptions were considered as an option for smoothing the interface, a technique developed for III-V quantum well growth. Growth interrupts are generally avoided in Si MBE due to the background impurity uptake, particularly O and C. These

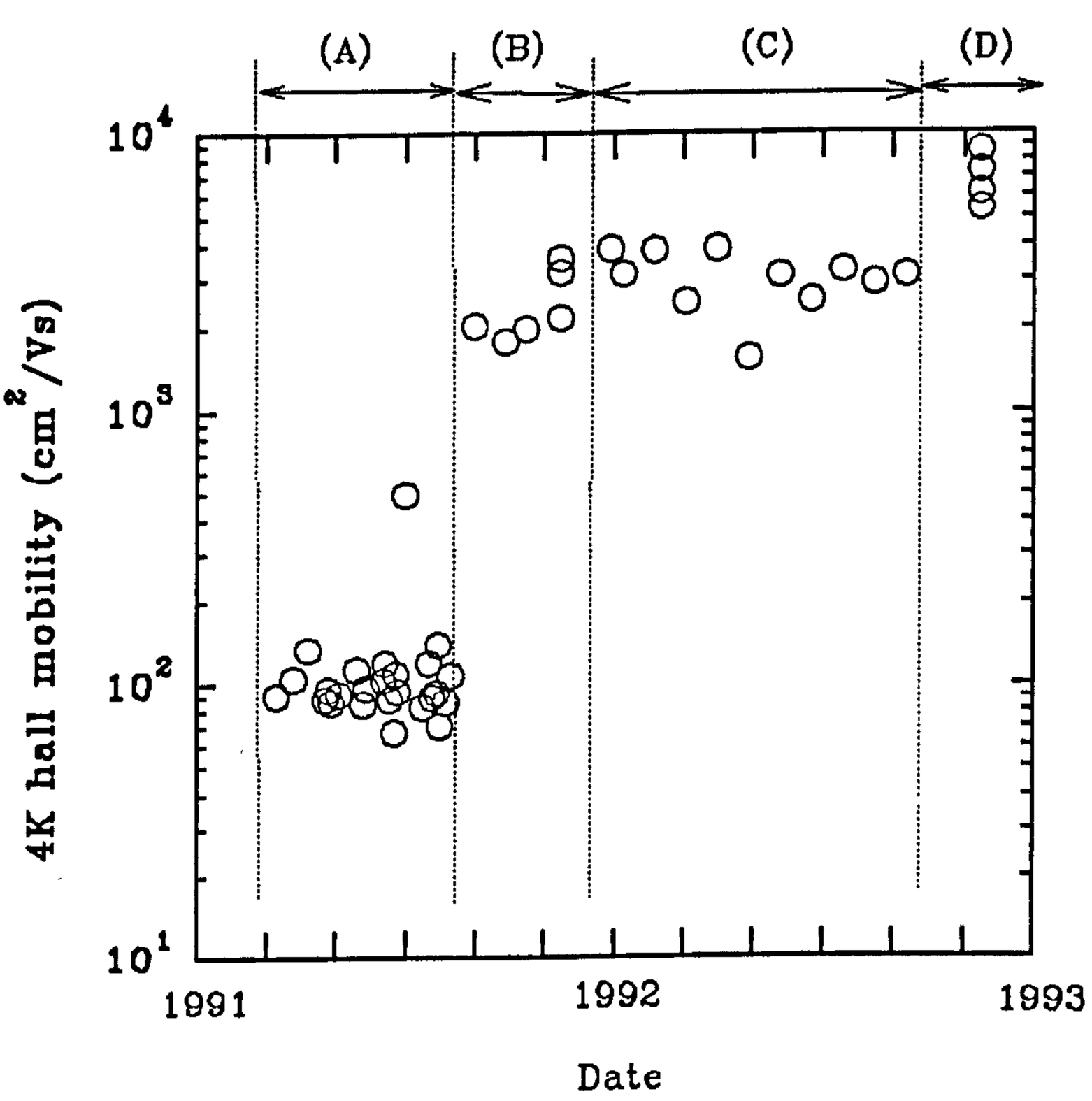


Fig. 5.10 Plot of Hall mobility at 4K for normal 2DHG's structures grown in the V90S system over two years. Each region (indicated by (A), (B), (C) and (D)) represents an alteration of growth conditions.

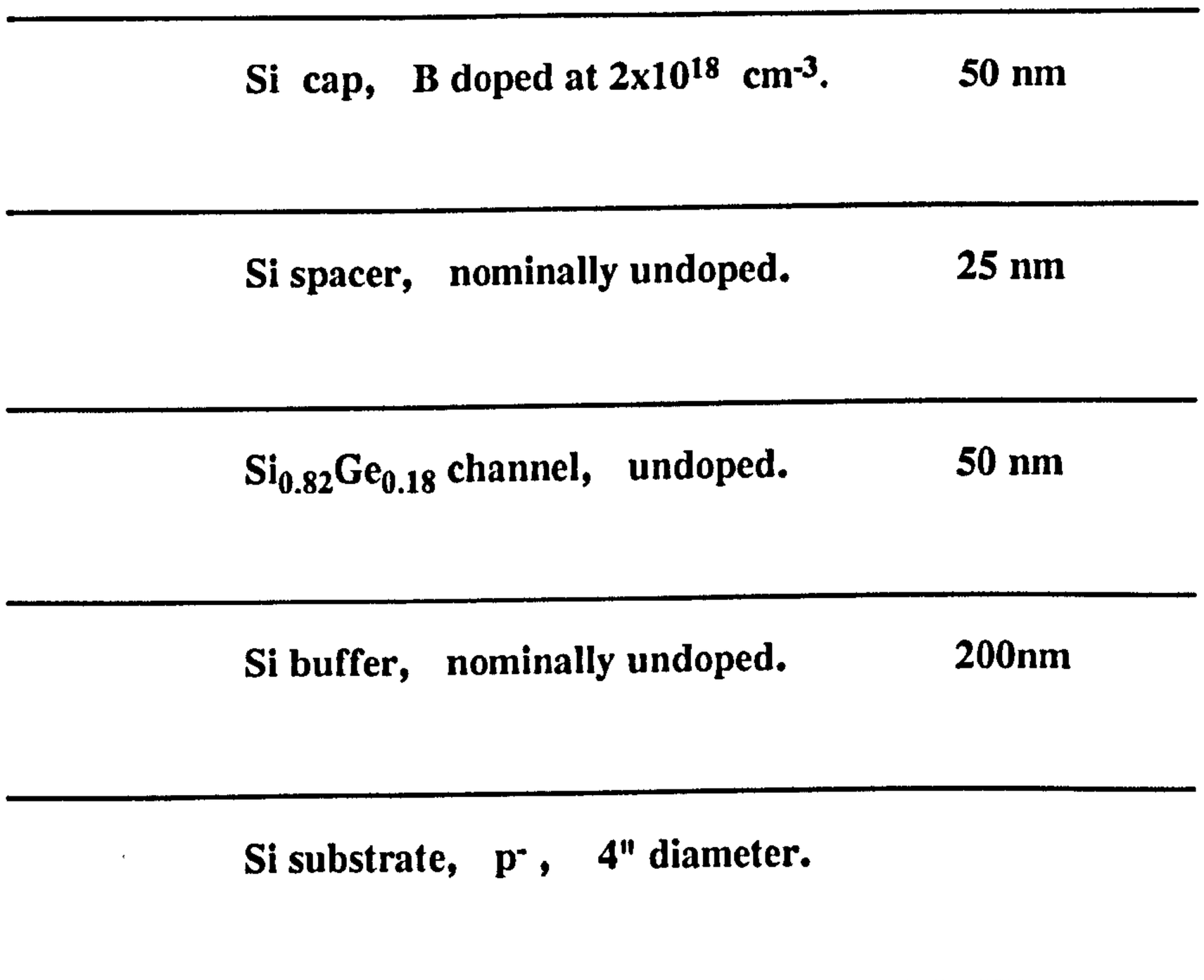


Fig. 5.11 Schematic of the 'normal' SiGe channel 2DHG structures used in this study.

interruptions aimed to provide further time for adatom ordering at the interface. Samples grown with interruptions were compared with an otherwise identical structure grown without an interrupt (sample A). Two three minute interruptions were used, one at or 0.5nm before initiation of the B doping (sample B), the other immediately on cessation of SiGe growth (sample C). T_s was 550^{0}C for all samples grown up to this point. The sample with the interrupt at cessation of Ge doping (sample C) had low temperature behaviour of mobility and resistivity that was identical to that of sample A, with no interrupt.

For sample A (no growth interrupt), the 4K mobility is low and is comparable with that expected from bulk doping of ~10^{18} cm^{-3} (Fig 5.12). The ρ versus T dependence suggests that states at the band edge are strongly localised. The cause of this localisation became clear subsequently. Plots of log G (ρ,T) versus T (see Section 5.2.1.1) for sample A (Fig. 5.13) yield a power of 0.35 $\pm$ 0.05 over 10-20K, indicative of 2D variable range hopping. Below 10K, the power decreases, the sample resistance, however, is approaching the noise limit for the measurement system, so this behaviour was not considered in detail. The mobility dependence of sample B, on the other hand, is indicative of conduction in a degenerate 2D gas with no strong localisation (μ[4K]) = 500cm^2 v^{-1} s^{-1}). The interrupt has produced a beneficial effect.

Close examination of the μ-T plot for sample B, Fig 5.12, reveals a minimum at about 15K. A similar minimum had been attributed in a GaAs/GaAlAs system to the scattering of a degenerate 2D gas of carriers by interface roughness, Sakaki et al 1987. However, cross sectional TEM of similar SiGe structures grown at Warwick had revealed abrupt interfaces to a resolution of 1nm: which is not consistent with this low mobility. One possibility considered, was that the minimum arose from resonant neutral impurity scattering (NIS) from an unintentional contaminant. The search for this contaminant was undertaken using Secondary Ion Mass Spectrometry (SIMS), for which the contaminant species must be identified.

A candidate for unintentional impurity in the SiGe channel is Cu, the e-beam hearth material. SIMS profiles of the two samples, obtained at 45° incidence, 2.0 keV

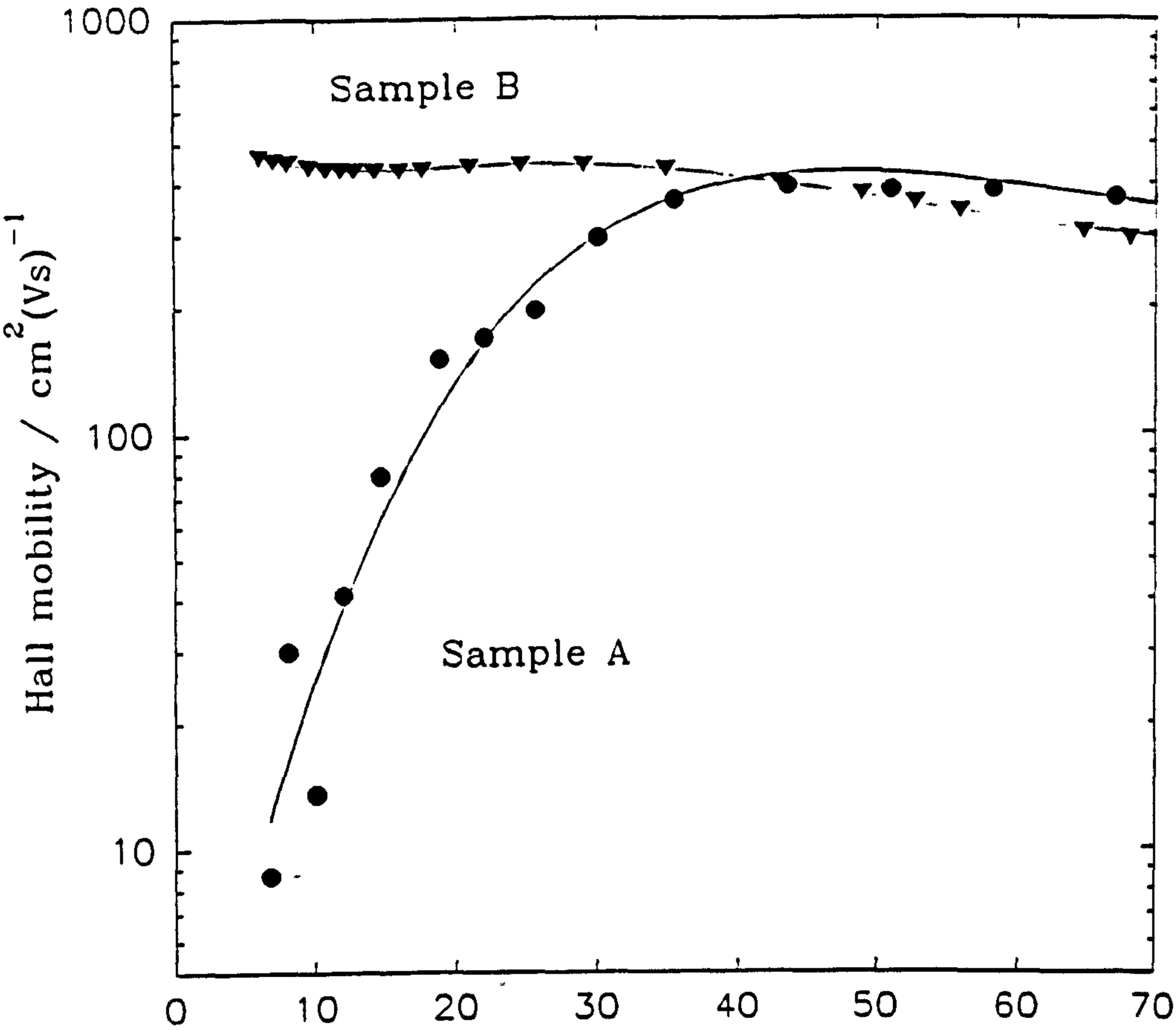


Fig. 5.12 Mobility versus temperature plots for 2DHG samples grown without (A) and with (B) growth interruptions.

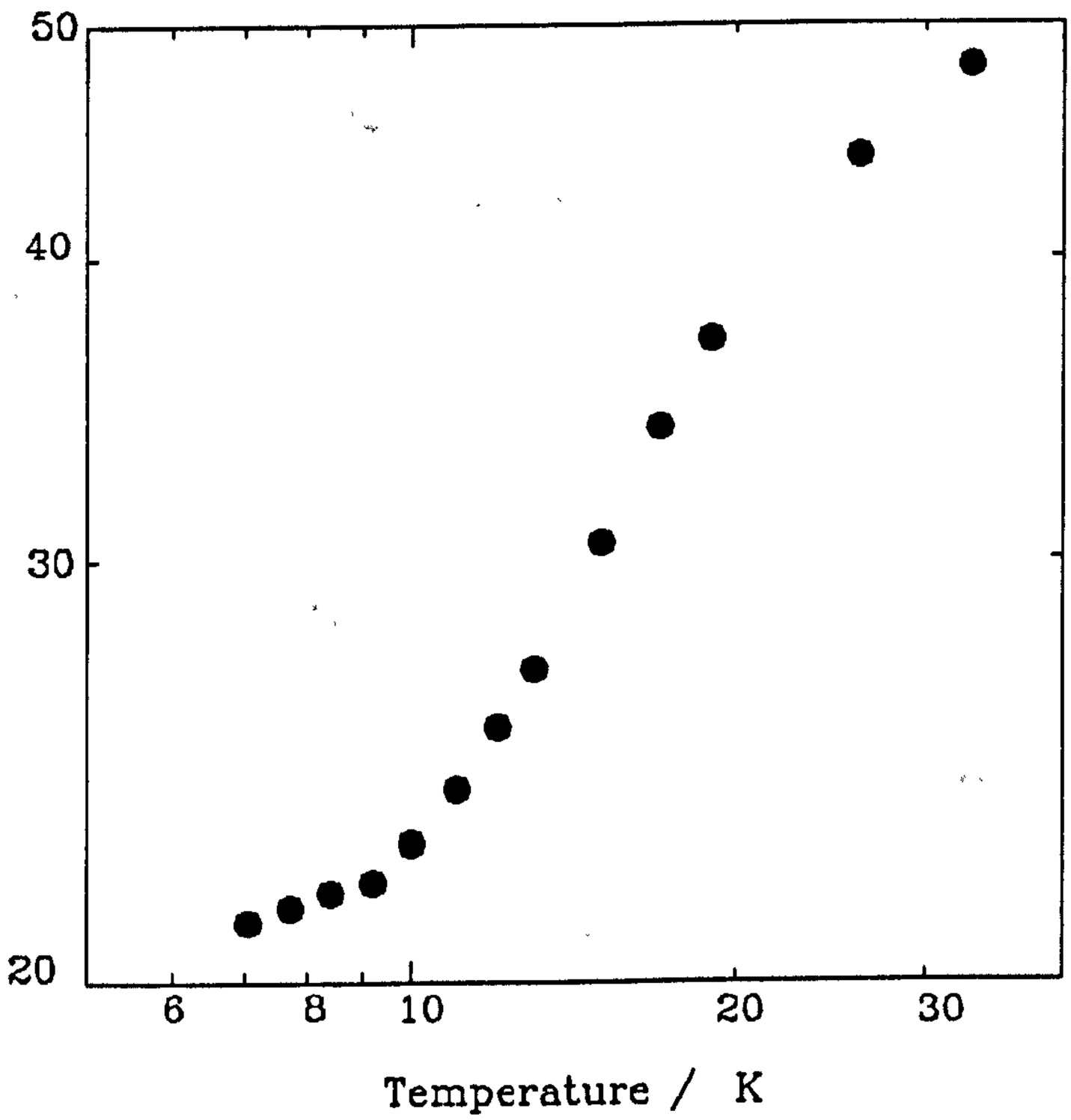


Fig. 5.13 Plot to assess localisation behaviour of sample A

are shown in Fig 5.14. This analysis demonstrates that the Cu concentration is very high - the peak level shown is greater than the solid solubility concentration in Si (Hall and Racette 1964). As such the absolute level cannot be considered accurate due to the Cu ion yield alteration introduced by precipitates. Further to this calibration problem, no standard for Cu in $Si_{0.8}Ge_{0.2}$ was available, so a Cu implant in Si was used. The most striking feature of the profile is that the Cu is sited in the SiGe for sample B, but, for sample A, resides mainly in the B doped Si region. Despite the SIMS calibration problems, this is considered to be a reliable observation.

It was postulated that the beneficial effect of the interrupt is produced by diffusion of metallics to a region of disorder introduced by the interruption, and subsequent retention of the metals in the B doped material, which acts as a preferential gettering domain. Such a gettering process raises many questions and is discussed below. It was further suggested that the mobility in sample A is limited by residual metallics and that the previously observed strong localisation arises from Cu precipitation in the well.

5.4.3 Cu Redistribution

This section relates to the latter stages of phase A (see Fig. 5.10) and the early stage of phase B. The source of the Cu contamination was considered likely to be the e-beam evaporator hearth. At typical Ge deposition rates (0.01-0.06 nm s^{-1}), the entire surface of the Ge charge melts, leading to inevitable contact between molten Ge and the Cu hearth. By interposing a Si liner, which remains solid at Ge evaporation temperatures, around the Ge charge, Cu dissolution into the Ge melt is eliminated. (but see page 117) The 2DHG mobilities of samples identical to A and B, but grown with a Si liner, in conjunction with the growth interruption, increased four-fold from 500 to 2200 $cm^2V^{-1}s^{-1}$. Importantly, these values were reproducible. This reproducibility is indicated in Fig. 5.10. After growth of a structure with a beneficial interrupt (the 500 $cm^2V^{-1}s^{-1}$ mobility of Phase A), further structures were grown

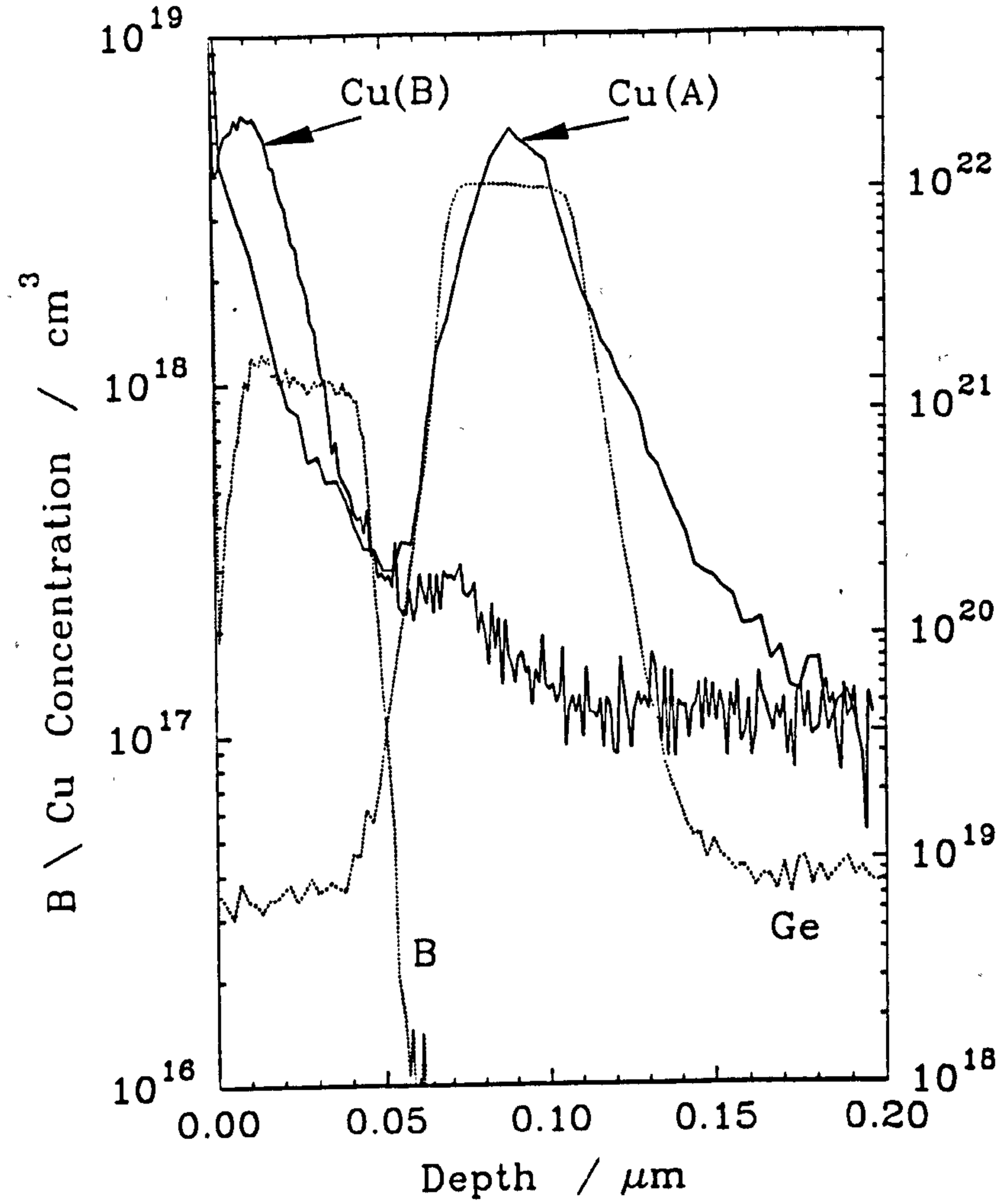


Fig. 5.14 SIMS profiles of 2DHG structures grown without (A) a growth interruption and with (B) a growth interruption. B and Ge concentrations are shown for one sample only to aid clarity.

without interrupts and with an unscreened Ge source, and 4K mobilities were consistently ~100 $cm^2V^{-1}s^{-1}$. Subsequently, (phase B), structures grown with interruptions, and a Si liner to the Ge, yielded 4K mobilities consistently ~2000 $cm^2V^{-1}s^{-1}$. Note that, in contrast, irreproducibility had been previously reported as a problem with the growth of similar structures (Van Gorkum 1992). SIMS on these high mobility samples indicated Cu levels below the detection limit ($\approx 1x10^{17}$ cm^{-3}), throughout the structure. It is conceivable that the Cu redistribution phenomenon is concentration dependent, related to Cu precipitation or is specific to fast diffusers in Si such as Cu. However, the improved reproducibility suggested that it may be effective with lower metal concentrations. This pre-supposes that the mobility is limited by Cu when using a screened Ge source: this argument is developed in section 5.4.5.

The apparent gettering of the growth interrupt is surprising. Following growth of the channel, the sample is at T_s for a further 10 to 15 minutes for the Si cap growth. Using a Fickian model for Cu diffusion predicts $\sqrt{Dt}$ for such a case as ≈850 μm; by comparison, the 4" substrate is 450 μm thick, the total epilayer is 2 μm thick. From the SIMS of sample A in Figure 5.14, the Cu from the Ge source is trapped in the SiGe region, presumably by strain. An interrupt could be expected to introduce some local disorder due to C and O related defect nucleation. Cu diffuses in Si via an interstitial mechanism. (Transition metal diffusion in Si is comprehensively reviewed by Weber, 1980.) Cu diffuses in Ge via an interchange between Cu interstitials and substitutional Cu, assisted by vacancies - a mechanism discussed by Stolwijk et al, 1984. While a high concentration of vacancies might be introduced as a result of the growth interruption, given the fast diffusion of Cu in nominally undefected Si, it is not clear why this should have such a significant effect. Further, this would not, in itself, explain why the Cu diffuses preferentially to the B region and is retained (from comparison of the area under the SIMS profiles ~95% of the Cu is redistributed). B at 10^{18} cm^{-3} introduces strain in Si, but not equivalent to that of alloy Ge concentrations. One could speculate that the O (or C) uptake at the interrupt, and possible formation of SiO_2

microcrystals, is relevant, but in the absence of further experimental evidence, there is little value in such discussion.

We were unable to study this process by intentional co-doping with Cu because of the risk of unnecessary contamination of the growth system. To gain further evidence regarding this phenomenon, an attempt was made to reproduce the results, and to identify the key requirements for preferential gettering, by post growth contamination. Four SiGe channel structures grown with a Si liner about the Ge (i.e. with low Cu contamination) were used in a post growth contamination experiment. Two structures were equivalent to samples A and B (i.e. without and with an interrupt), the other two were repeats of these but with no B doping. These structures were chosen to yield information regarding the role of B in the Cu redistribution.

Cu was sputtered onto the surface of these structures for two minutes using a rotating carousel to minimise heating. Surface profilometer measurements using a Dektak gave a Cu thickness of 40 nm. Samples were then annealed in an RCA cleaned furnace at temperatures of 550 C and 640 C for 15 minutes to replicate the thermal budget of the Si cap growth. The subsequent diffusion profile is modelled to a first approximation as;

$C(x,t) = \frac{C_0}{\sqrt{\pi Dt}} e^{\left(\frac{-x^2}{4Dt}\right)}$, where C_0 is the surface concentration.

At the lowest anneal temperature, 550 C, the concentration predicted at the lower SiGe interface, C(125 nm, 900s), is found to be less than 10^7 cm^{-3}.

SIMS analysis revealed no preferential siting of the Cu at either anneal temperature in any of the structures. Indeed, the Cu had diffused such that it was below the SIMS detection limit throughout the active region of the structure (the SiGe channel), as would be expected for bulk Si. Fig 5.15 is two overlaid SIMS profiles of an annealed 2DHG structure with no intentional Cu contamination, (i.e. grown using a screened Ge source), and an annealed sputtered Cu contaminated sample. The profile of the uncontaminated structure reveals minimal B or Ge diffusion. The Cu contaminated sample is very different. There are two features of this profile; firstly, that the Cu has

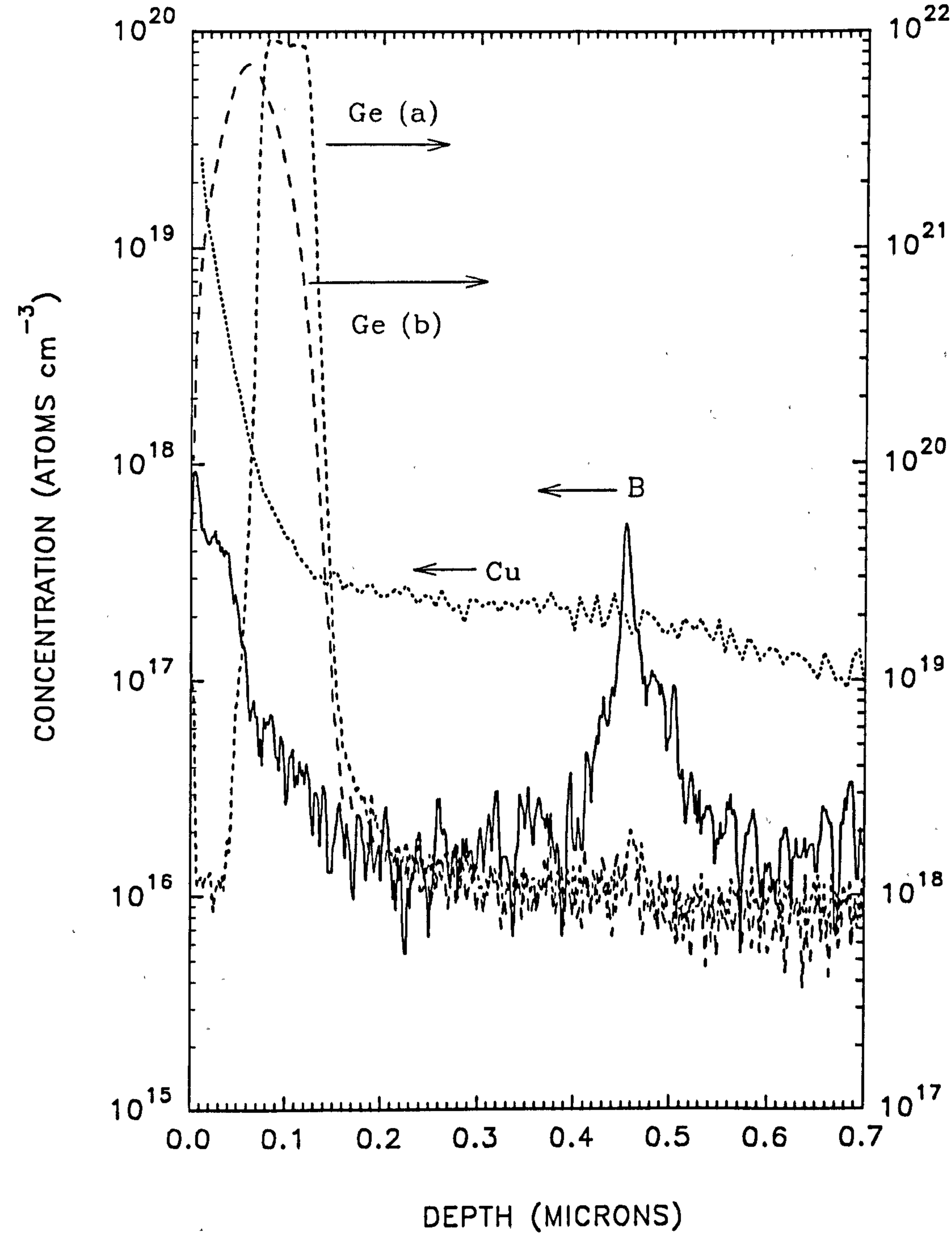


Fig. 5.15 Overlaid SIMS profiles of two 2DHG structures annealed at 640 C for 15 minutes: (a) had no intentional contamination, (b) was sputtered with Cu. Profile conditions are 4KeV, O^{2+} at normal incidence for Ge and B , 45° for Cu. (SIMS by R. Barlow, Warwick.)

diffused and is not preferentially sited in the Ge or B region and, secondly, that the Ge appears to be very close to the surface and broadened. It is not yet known why this latter effect is seen - it was common to all Cu contaminated annealed samples. There is currently a lively debate in the SIMS community about the effect of dislocations on Ge profiles: the author has no intention of joining this debate. These structures were heavily dislocated after annealing, so it is possible that SiGe relaxation on annealing, due to Cu contamination, plays a role in this result.

It is also interesting to note the conclusions reached by Higgs et al 1990, who found that intentional Cu contamination greatly enhanced dislocation formation during SiGe annealing. They suggest that previously reported order of magnitude variations in experimental values of metastable t_c, the critical thickness, may be a result of metal contamination. The SiGe used for these t_c studies was grown by MBE, but not at Warwick. Indeed, further measures taken with the aim of reducing residual metal contamination, as will be described, increased the 4K 2DHG mobility to ~$4x10^3$ $cm^2V^{-1}s^{-1}$ in otherwise identical structures. Metallic contamination might possibly be a universal limitation in MBE Si.

Clearly, the Cu redistribution has not been repeated by this experiment, thus, it must be concluded that it is a growth related phenomenon, and probably cannot be elucidated or studied by post-growth experiments. It is interesting to note that similar Cu redistribution resulting from growth interruptions in MBE SiGe was reported by Ni, 1992. Unfortunately, no further details of this observation are available.

5.4.4 T_s Dependence

This and the following section relates to the latter stages of phase B (Fig. 5.10). Using growth interruptions with a Si liner about the Ge source, structures were grown with 20nm spacer, Ge fraction 0.18 and n_s ~ 2 x 10^{11} cm^{-2}. T_s was varied between 520 C and 640 C. The results are shown in Fig 5.16. The increase in mobility with T_s

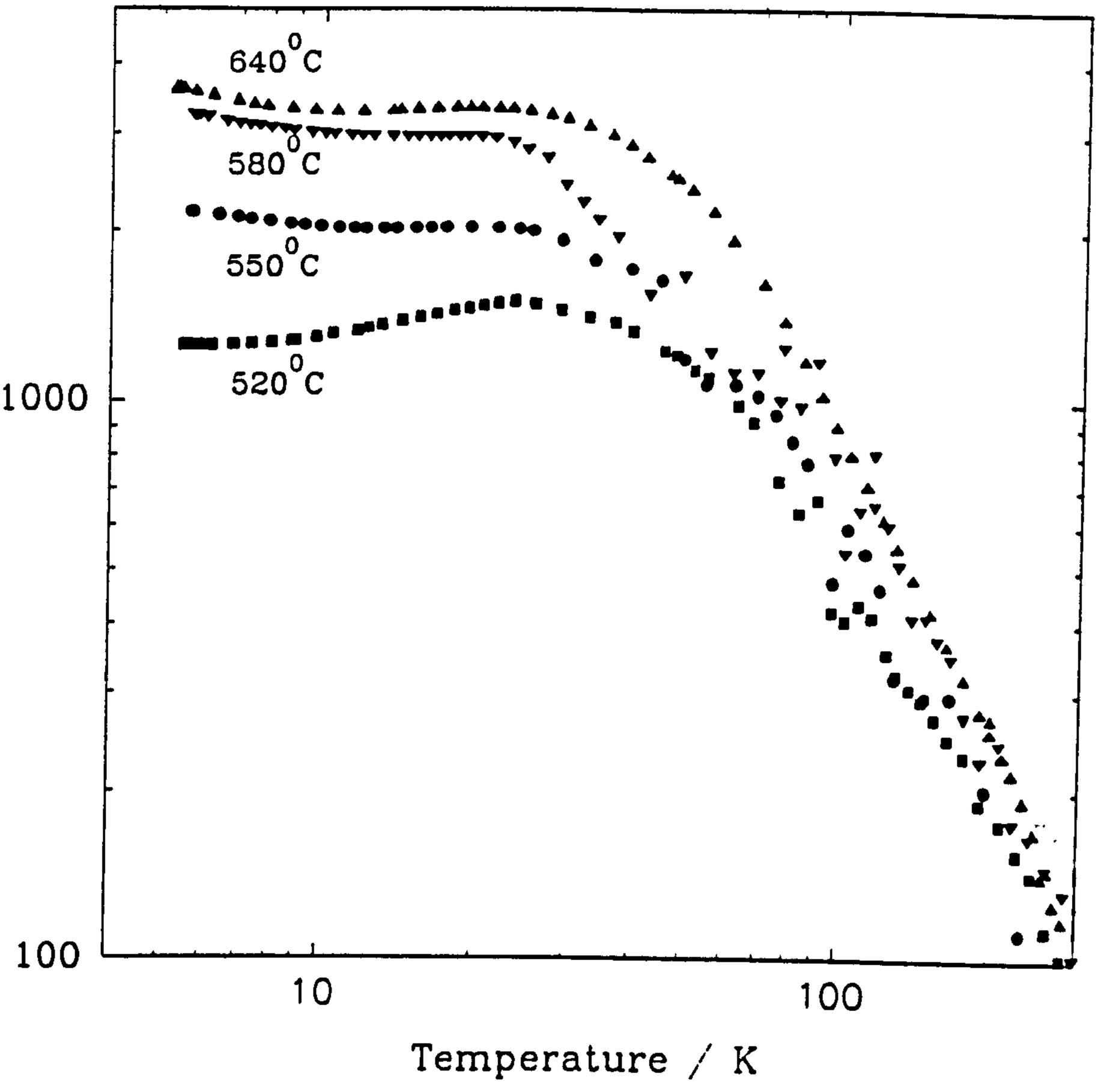


Fig. 5.16 Variation of mobility with Ts for 2DHG's in a $Si_{0.82}Ge_{0.18}$ channel with sheet concentrations $\approx 2 \times 10^{11}$ cm^{-2}.

is the most striking feature of these results with a maximum $\mu = 3640\ cm^2\ v^{-1}s^{-1}$ at 5.4K for $T_s = 640$ C. Note that minima exist in all of these μ-T plots.

Care must be taken to consider fully the possible reasons for this result. It is conceivable that the mobility may be dominated by ionised impurity scattering, although Hess, 1979, calculates a very weak μ-T dependence only for this scattering in a degenerate gas, and no minimum is predicted. Ionised impurity scattering could arise from the background n-type impurities sited in the SiGe channel, or by remote doping from B in the spacer layer. The background n-type impurity concentration in the V90S system is $<1x10^{15}\ cm^{-3}$: the background n-type impurity in the V80 is phosphorus, which shows no discernible dependence on T_S in Si (Houghton, 1991), although background doping has not been studied for SiGe. In addition, B diffusion into the spacer layer, which would degrade the mobility, will increase with increasing T_S. B segregation effects, which do reduce with increasing T_S over this temperature range (Parry 1991), are not pertinent for normal structures other than to increase the effective spacer layer thickness. An important phenomenon reported to have the correct temperature dependence is Ge segregation. Nakagawa and Miyao, 1991, assert that Ge on Si segregation reduces with increasing T_S over the range of this study. However, there is no clear reported evidence for significant segregation in the present, relatively dilute $Si_{0.8}Ge_{0.2}$/Si system: such a segregation dependence was not, moreover, identified by X-ray analysis in similar samples. Increasing T_S, however, may act to smooth the SiGe/Si interface, thereby relieving interface roughness scattering (Mori and Ando 1980). Cross sectional TEM of this sample set reveals abrupt interfaces to a resolution of 1nm, as mentioned previously. None of these foregoing explanations of the T_s dependence is satisfactory: in the following section, interface charge, possibly arising from metal contamination, is proposed as the limiting scattering mechanism.

Post growth annealing can be expected to yield information about whether the Ts dependence is a growth related or solid state diffusion process. Samples grown at T_s=550 C were annealed at 640 C for 15 mins in an RCA cleaned tube furnace in a flowing N_2 ambient. The results are shown in Fig 5.17. The 4K mobility is unchanged

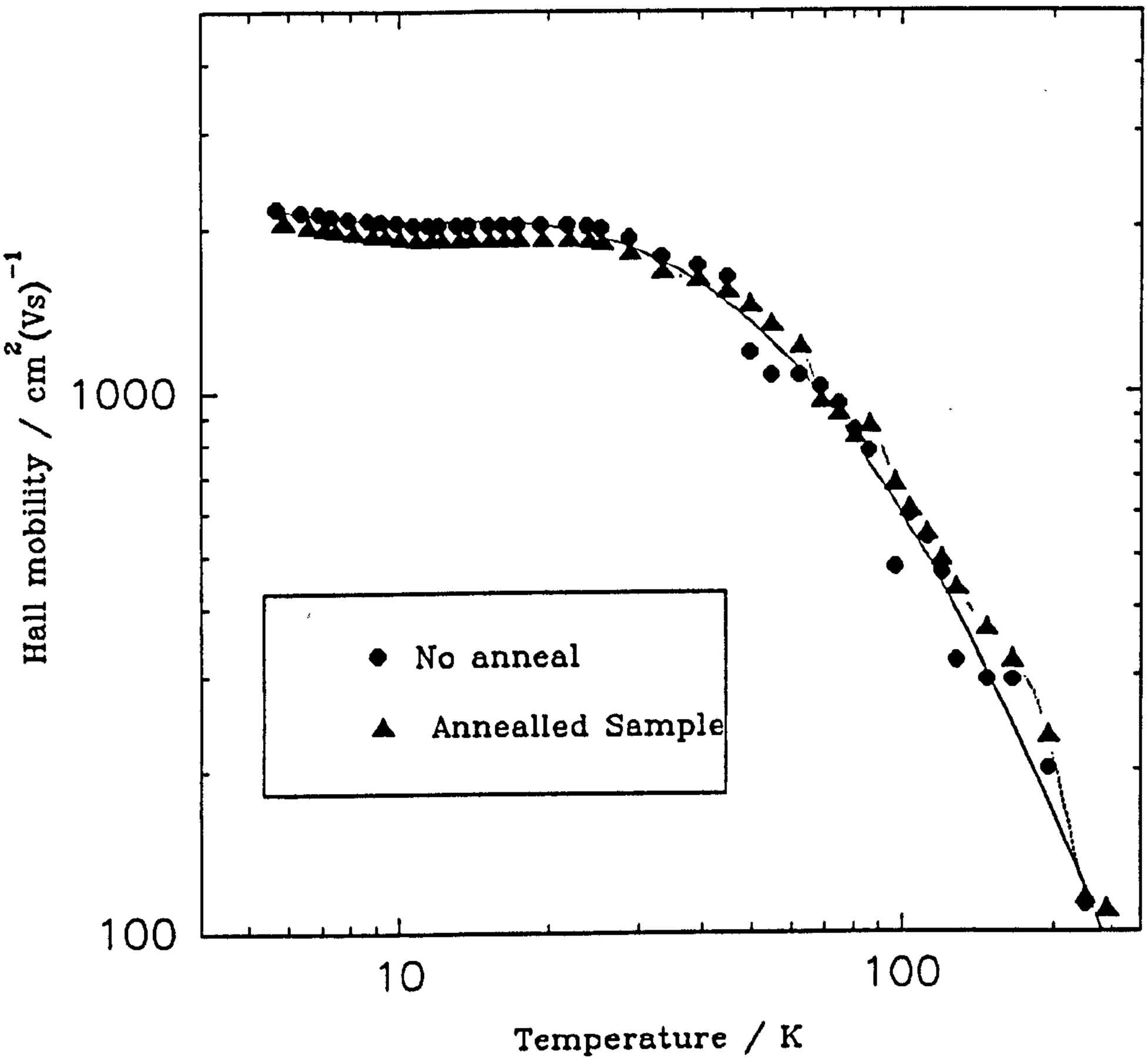


Fig. 5.17 Mobility versus temperature plots from a 2DHG structure (T_s=550 C) before and after annealing at 640 C for 15 minutes

unchanged within experimental error. Therefore, the T_s dependence is seen to be associated with growth processes.

5.4.5 Limiting Scattering Processes

The structures grown in the aforementioned investigation of T_s dependence were used in low temperature parallel transport studies (by the author and co-workers) and capacitance voltage (CV) profiling. It is useful, at this point, to give brief summaries of three studies, insofar as they are relevant to the growth studies of this chapter.

The first study (Emeleus et al 1992) considers the low temperature (0.3K - 20K) behaviour of the conductivity, σ_{xx}, and Hall coefficient. Quantitative analysis reveals that conduction is governed by both weak localisation and hole-hole scattering and that the decrease of σ_{xx} between 2 and 6K is a temperature dependent screening effect, explicable in terms of theory developed by Gold and Dolgopolov, 1986. This is the origin of the minimum observed in the μ-T plots of Fig. 5.16. These effects were observed in samples with μ(4K) both of 2,000 and 3,640 $cm^2V^{-1}s^{-1}$, i.e. grown at 550 C and 640 C.

It was postulated by the author (Smith et al 1992), that the T_s dependence of mobility could be associated with residual metal contamination at the Si/SiGe interface. Contamination of this type can be modelled as charge at the interface: negative charge at the interface will induce band bending that decreases the width of the quantum well. Thus, the carrier concentration will be larger than otherwise predicted. The second study summarised (Emeleus et al 1993) relates to n_s and calculations of μ. The model for n_s developed in section 5.4.1 is readily modified to account for a sheet of charged impurities, n_i, at the interface by including it in the expression for the field at the interface, F_0

$F_l = \frac{\varepsilon}{\varepsilon_0 \varepsilon_r}(N_A l + n_i)$,where l is the depletion width in the B doped region.

The sum of the potential variations in the well, to the top of E_V, is then compared with experimentally determined band offsets. The interrelation between well shape and n_s is approximated by assuming a spatial wave function of the Fang Howard type (Ando et al, 1982). The well shape resulting is found from Poisson's equation by using the usual expression for energy levels in a triangular well, but replacing F by an effective field, F_{eff} defined by $E_0(F_{eff})=eaF_{eff}$, where a is a variable well width parameter. This allows self consistent solutions for n_s to be obtained as a function of n_i. The value of n_i required for good agreement with experiment is negative charge with a sheet concentration of $2x10^{11}$ cm^{-2} (Fig 5.18). This is high, but compares with the value found necessary to fit the 4K mobility of 2000 $cm^2V^{-1}s^{-1}$ to calculated values, assuming Coulomb scattering from charged interface impurities, as described by Gold and Dolgopolov, 1986.

In the third work (Brighten et al 1992), CV measurements of the heterojunction were compared to theory and gave strong evidence for charge at the interface with sheet density ~ 10^{11} cm^{-2}. The temperature dependence of these measurements indicates acceptor-like, negative, charge.

It must be noted that this evidence relates to negative charge at the Si/SiGe interface and that Cu is a triplet acceptor state in Si and Ge. Also note that our original hypothesis of neutral impurity scattering is not necessary to explain any feature of transport in these structures. This hypothesis, however, proved valuable because it led to the investigation of unintentional contamination reported here.

We, therefore, believe the T_s dependence to be related to a reduction in interface charge density. The scattering mechanism as determined by low temperature parallel transport studies is common to all of the samples, although the strength of scattering obviously declines for high T_s samples (note that there is a minimum in μ versus T even for the highest mobility sample). This scattering mechanism is confirmed by the presence of deep acceptors at the interface as revealed by CV studies.

It is worth considering whether the interface charge, and thus the mobility, in these samples is associated with metal contamination. Although we cannot demonstrate

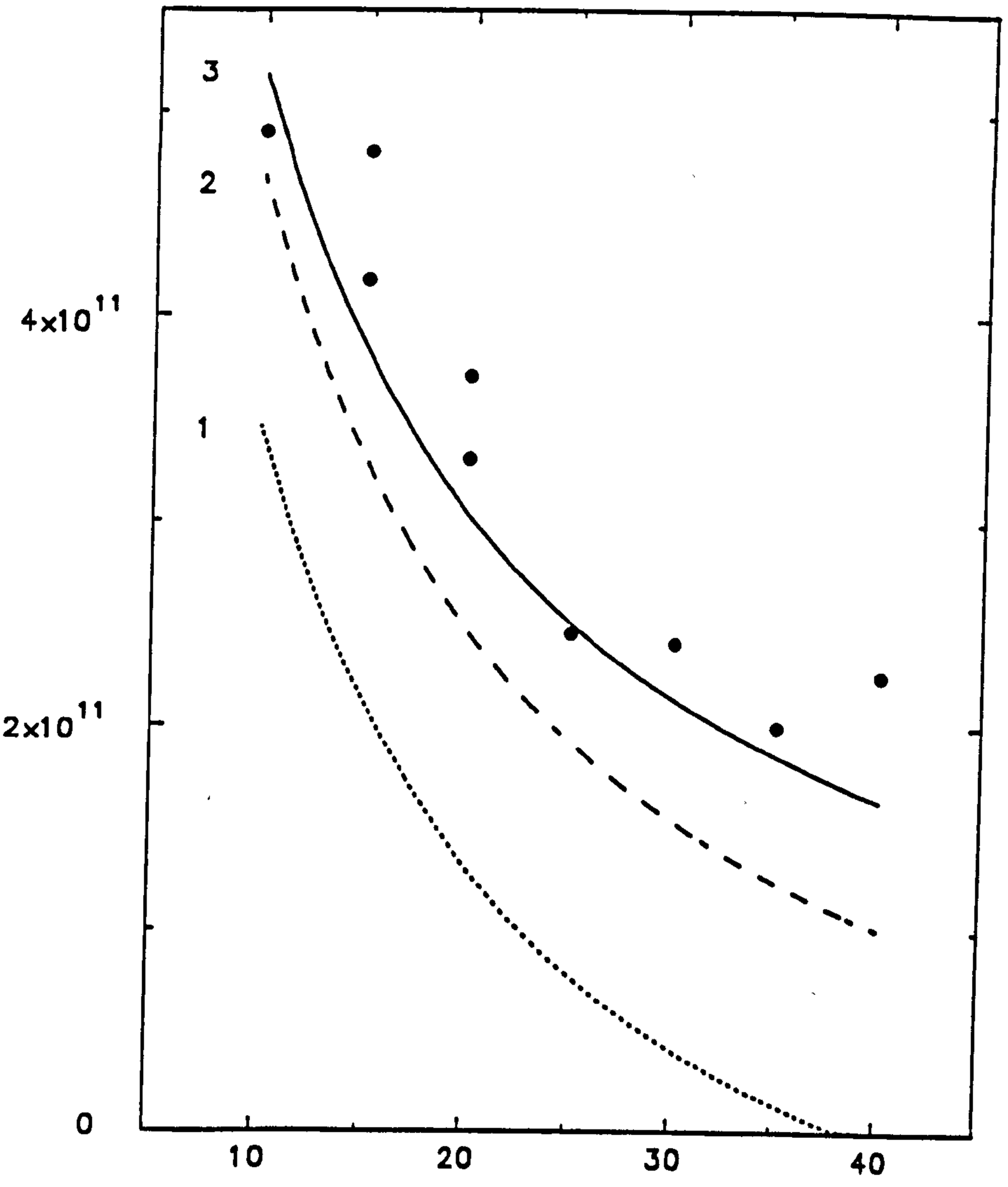


Fig. 5.18 Variation of sheet hole concentration with setback. The solid circles are experimental data, the lines represent self consistent calculations for various values of interface charge (upper curve is found for $n_i = 2 \times 10^{11}$ cm^{-2}, N(depletion) $= 1 \times 10^{11}$ cm^{-2}, middle for $n_i = 0$, N(depletion $= 7 \times 10^{10}$ cm^{-2}, lower for $n_i = 0$, N(depletion $= 1.7 \times 10^{11}$ cm^{-2}).
(After Emeleus et al 1992)

a clear correlation (e.g. through the use of SIMS), between the interface charge and metal contamination for the low Cu levels achieved with the use of a Si liner, this correlation is made for higher Cu concentrations. Moreover, the nature of the dominant scattering mechanism is common to all samples with mobilities > 500 $cm^2V^{-1}s^{-1}$ i.e. including those in which metal concentration has been directly observed. Interface charge, however, could arise from other sources of contamination or from charged defects.

5.4.6 Si Charge Liner

This section relates to phase C (see Fig. 5.10) of these growth studies. Cu is a fast diffuser in Si. Although the Cu hearth is water cooled and the Si melt is contained within solid Si, the hearth has been estimated to reach 80°C during growth. This leads to a diffusion rate of 0.8 $\mu m\ s^{-1}$ i.e. Cu will diffuse to the centre of the charge in ~ 60 hours use. Thus, Cu contamination can be expected during the life of a Si charge, which can be up to 8 weeks. This is exacerbated by Si charge depletion during this time and, of course, Cu, when in the Si charge, will be subject to higher temperatures. Whilst Cu contamination resulting from the Si source will not be as extreme as that found from the use of an unscreened Ge source, it may be significant in these 2DHG structures. Evidence for this, and the continued role of metals in these structures, came from the introduction of a graphite liner to the Si charge. This had the effect of increasing μ(4K) at T_s = 550 C from 2,200 to 3,800 $cm^2v^{-1}s^{-1}$. This discussion also relates to the use of a 0.5 cm thick Si liner around the Ge charge, which will rapidly become contaminated; as a result, this was also replaced by a graphite liner.

5.4.7 High T_s Dependence

This section relates to phase D (see Fig. 5.10) of these growth studies. A further investigation of the mobility dependence on T_s was made using the graphite

shielded charges. Previously, T_s had been restricted to 640 C, the highest T_s believed compatible with the retention of 2D growth processes. This investigation was extended to T_s = 900 C and the results are shown in Fig 5.19. In order to try to prevent 3D islanding at high growth temperatures, the Ge fraction, x, was reduced to 0.13. In addition, since dislocation movement is an activated process, reducing x will reduce strain and therefore might reduce the dislocation density at high T_s. This helps to prevent relaxation and thus to eliminate dislocation density as a variable. The temperature dependence of Fig. 5.19 can be seen to follow a similar trend to that of the previous study. The surprising result is that μ (4K) continues to increase with T_s up to 850 C. The evidence that this trend might be related to metal contamination is discussed later in this section. Firstly, we shall discuss the saturation of μ with T_s and the decrease for $T_s > 850$ C.

The Si/SiGe interface roughness is characterised by two parameters - Λ, the roughness perpendicular to the interface and λ, the in-plane roughness correlation length. The calculations of scattering strength for interface roughness previously cited (Emeleus et al 1992) were performed for $k_f\lambda = 1$, i.e a worst case resonant condition of λ=10nm The value of Λ used is ±0.5nm as measured in structures grown below 800 C. These calculations predicted μ(4K) of between 1 and 2 x10^4 $cm^2V^{-1}s^{-1}$ for these values of n_s.; thus, interface roughness scattering has been discounted previously. (Note that there is no evidence for λ=10nm in these structures, so the interface roughness limited mobilities for $T_s <$ 850 C are expected to be $> 10^6$ $cm^2v^{-1}s^{-1}$.) Cross sectional TEM of samples grown at 800 C and 900 C demonstrate vertical roughness of <1 nm and 10 nm respectively (Fig. 5.20). Even at the highest T_s there is no clear evidence for relaxation, but there is evidence of long range SiGe channel thickness variations analagous to the ripple morphology seen in VPE growth (Pidduck et al 1993). If values of Λ and λ, taken from the TEM of the T_s = 900 C structure, are used, then the theory (Gold and Dolgolopov, 1986) predicts $\mu \sim 10^4$ $cm^2v^{-1}s^{-1}$. Approximate agreement between absolute values of mobility and calculations is strong evidence for

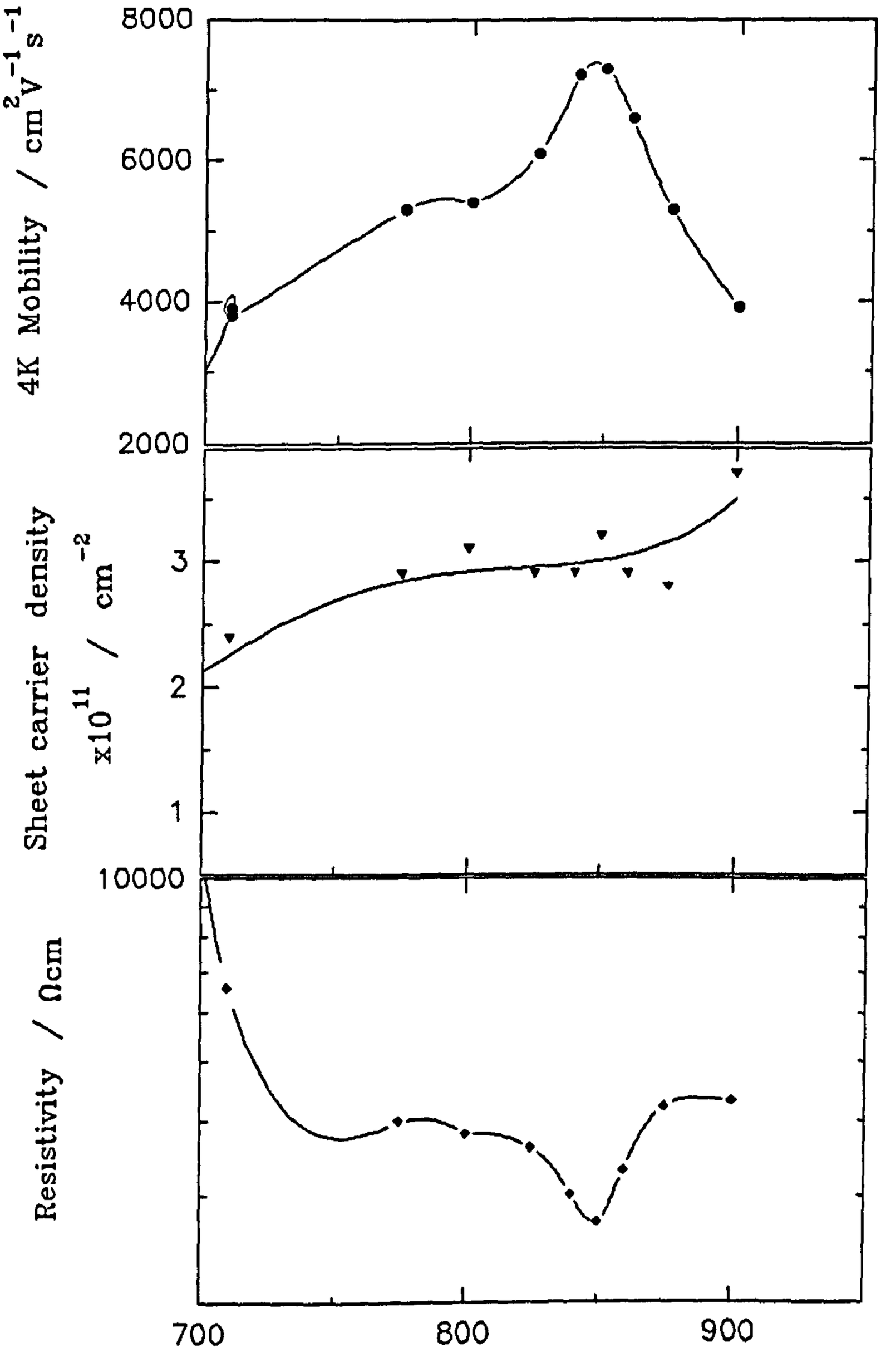


Fig. 5.19 Measured parameters of 2DHG's in $Si_{0.87}Ge_{0.13}$ channel, remote doped structures grown at high T_s

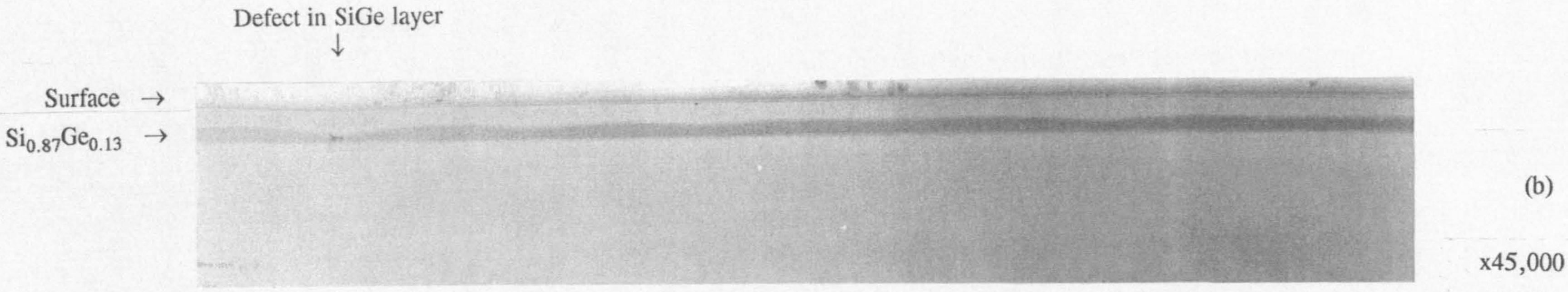


Fig. 5.20 TEM micrographs of 'normal' remote doped 2DHG $Si_{0.87}Ge_{0.13}$ channel structures (SiGe thickness is 50 nm). The structure shown in (a) was grown at T_s=700 C, that in (b) at T_s=900 C.
(Analysis by P. Augustus)

interface roughness scattering, but is not in itself convincing, so further investigations were made.

The relation between μ and n_s for samples grown at 900 C, was investigated by varying the undoped spacer layer width in the range 2.5 to 30 nm, Fig. 5.21. It can be seen that the mobility falls for both high and low n_s. The decrease of mobility for high n_s is in accordance with the trend expected from interface roughness scattering (see Section 5.2.2). This is further evidence that the SiGe ripple can limit mobilities. The decrease at low n_s, however, suggests that interface charge scattering remains significant, even though the mobilities are consistent with a much lower interface charge density.

Information regarding the scattering can be provided by decreasing the SiGe channel width, which can broaden the subbands in the presence of interface roughness. The spatial extent of the wavefunction for $n_s = 2.5 \times 10^{11}$ cm^{-2} was studied by reducing the SiGe channel thickness in a series of samples, Fig 5.22. There is a dramatic fall in μ(4.2K) for channel widths, a, less than 20 nm. The well width for $n_s = 2 \times 10^{11}$ cm^{-2} is less than ~7nm, so the 2DHG cannot be assumed to be scattering from both lower and upper interfaces. A similar fall in μ with channel thickness has been previously observed in III-V systems by Sakaki et al, 1987, in the presence of interface roughness scattering and was shown to have a $\mu \propto a^6$ dependence. While there is insufficient data from which to confirm any relation in this study, it is noted that the data is more consistent with the relation $\mu \propto a^2$. This suggests that interface charge scattering is also present, in agreement with the interpretation of the μ versus n_s behaviour. It is therefore suggested that the increase in mobility with growth temperature is associated with a reduction in interface charge scattering, and that the mobilities for $T_s > 850$ C are limited by both interface charge and interface roughness scattering. This would appear to be the most promising explanation of the saturation of mobility with high T_s

At this point, it is worth summarising the argument that the T_s dependence of μ (4K) may be metal related. Cu has been directly observed in samples with the bulk-like 4K mobilities, and has been deduced to have greatly reduced concentration at the

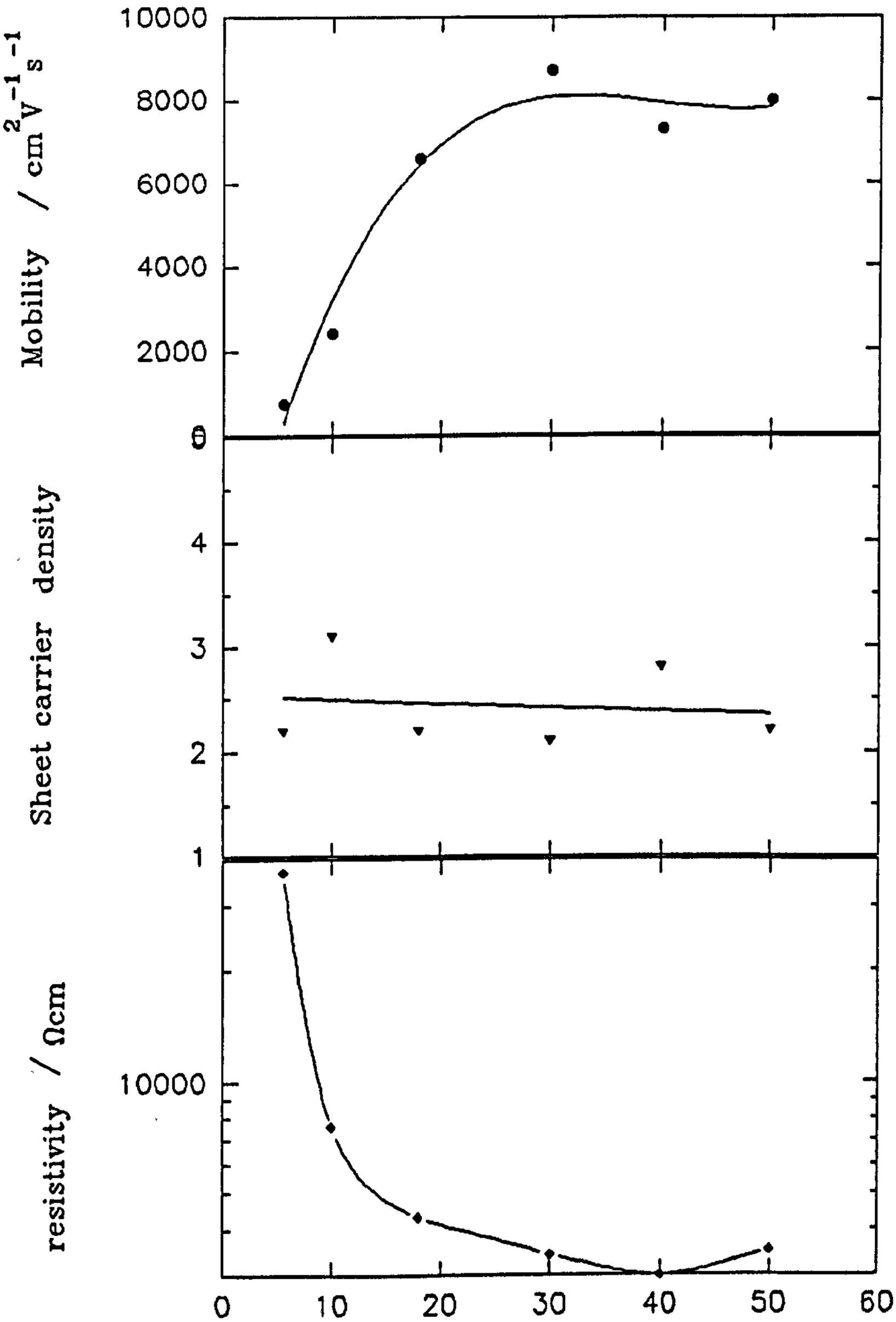


Fig. 5.21 Measured 2DHG parameters as a function of spacer thickness. ($Si_{0.87}Ge_{0.13}$ channel).

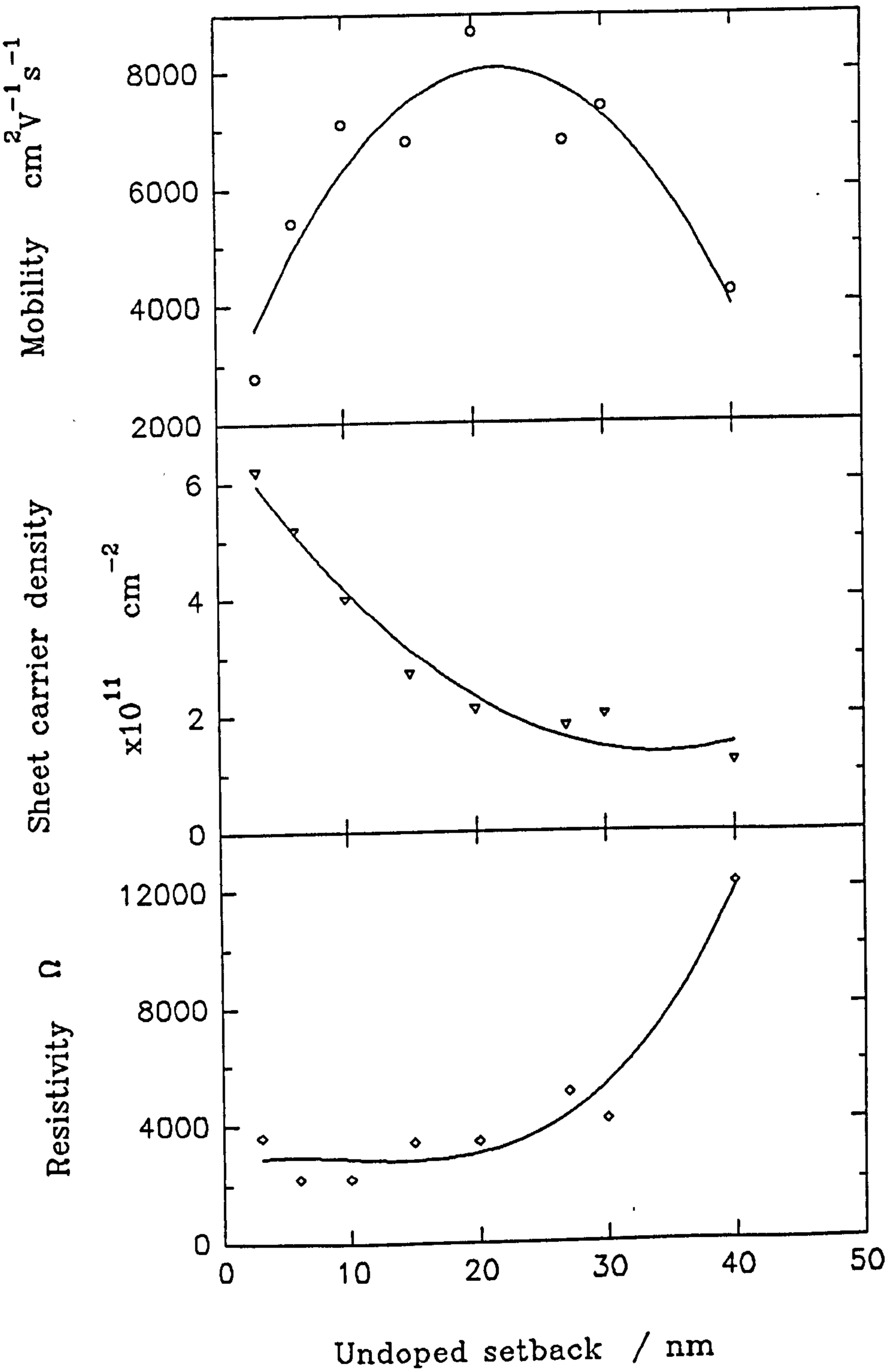


Fig. 5.22 Measured 2DHG parameters as a function of $Si_{0.83}Ge_{0.13}$ channel width

confining interface in samples with higher mobilities. The limiting scattering mechanism is common to samples grown at T_s = 500 C and at T_s = 640 C with μ in the range 2000 - 4000 $cm^2V^{-1}s^{-1}$. Acceptor-like charge has been observed at the SiGe/Si interface of samples grown under identical conditions. Experimental values of carrier concentration in the well can only be fitted to a self-consistent model assuming acceptor-like interface charge. Measures taken to reduce Cu contamination in the source charges enhanced μ(4K). The only evidence linking the gettering effect of the growth interruption to samples with Cu concentration lower than the SIMS detection limit is the improved reproducibility encountered when using the interrupt. The role of the interrupt must therefore remain speculative.

The second sample set with T_s as variable parameter (i.e. 700 C to 900 C) exhibits a similar T_s dependence, for T_s < 850 C, to that of the first. There is not yet any magnetoresistance, R_H analysis linking the two sample sets other than the evidence for continued interface charge scattering and interface roughness scattering for T_s>850 C. It seems unlikely that there are two independent mechanisms both yielding the T_s dependence we have observed. Scattering calculations suggest that no other mechanism can limit the mobility to values below 10^4 $cm^2V^{-1}s^{-1}$, other than very large interface roughness which has not been observed by TEM. Therefore, it is proposed that this dependence may be related to interface charge, possibly from residual metal contamination, at the SiGe/Si interface.

CONCLUSIONS

Metal contamination has been implicated as the limiting factor for μ(4K) in SiGe channel remote doped structures grown with unscreened Si and Ge sources. A redistribution of Cu from the SiGe channel to a B doped region has been observed and is related to a growth interruption: this has been shown to be a growth related process. Modifications made to the growth system to reduce Cu contamination have enhanced the

mobility. A strong dependence of low temperature mobility on T_s has been observed in structures grown with Si and Ge source shielding. This has been identified as due to a reduction of interface charge by other workers. This effect has also been demonstrated, by post growth annealing, to be a growth related process. The combination of these measures has led to values of $\mu(4K)$ greater than twice as high as those previously reported.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

Si and SiGe grown by MBE can offer many benefits to the study of solid state physics, and may prove to be useful in the production of new devices. However, the effects on electrical material quality of the interrelated processes of substrate preparation, growth technology and growth parameters are poorly understood and not extensively studied. The comprehensive study of these effects - to which the present work is a contribution - is a long-term goal of the drive to realise MBE structures both for commercial application and for new studies in low dimensional semiconductor physics.

The objectives of this study were to measure room temperature electrical quality of as-grown Si as a function of growth temperature and to establish the requirements for high mobilities, at low temperature, in SiGe channel 2DHG's formed by remote doping. These objectives have been largely met. It was necessary to modify post-growth processing techniques in order to realise some of the measurements: the effects of this processing on material quality are also of interest.

Carrier lifetimes are a direct measure of material quality and are important device parameters. Few previous values of lifetimes in Si MBE have been reported and most have been measured in MBE Si following high temperature post growth processing, which is likely to alter material quality. Reported values have varied by many orders of magnitude. No systematic study of lifetime dependence on growth temperature has been previously reported.

For the first time, generation lifetimes have been measured in MBE grown Si by an MOS technique where all post growth temperatures have been kept below the

growth temperature, T_s. These are believed to reflect as-grown Si quality. A study on V80 grown Si (which was limited by what is believed to be oxide breakdown effects), has found τ_g ~ 200 ns. No dependence of τ_g on T_s could be confirmed by this study, due to the oxide breakdown effects. A similar investigation of V90S Si, yielded a strong and complex, but repeatable, τ_g versus T_s dependence. The maximum values of τ_g (obtained both at high T_s, 700 C, and low T_s, 550 C,) are ~2 μs, as compared to ~10 μs in control substrates. The T_s dependence is unaltered by post growth annealing, which suggests that this is not a thermal effect (e.g. solid state diffusion) i.e. it is growth related. Values of τ_g measured in pn diodes processed at a VLSI facility are, in one case, high (10 μs), in the others, generally low (1-150 ns). This latter sample set also revealed a minimum at T_s~600-650 C, but the control substrates yielded low lifetimes. This, and the large variation of τ_g obtained across each epilayer renders these results generally unsatisfactory. The processing and mask sets used in this study were designed by the author, to preserve the integrity of high resolution doping structures. These process steps are not well characterised and it is believed that the results obtained from pn diodes may be dominated by the effects of processing. It has been established that one of these process steps - a short, high temperature implantation activation anneal (15 s, 1000 C) - does not inherently degrade material quality, as reflected by generation lifetime.

Relatively low values of SiGe channel 2DHG mobility have been uniformly reported by many MBE (and UHV CVD) groups. The reason for this was unknown and it had been suggested that low mobilities might be inherent to the structure, due to alloy scattering. An extensive investigation of the material quality requirements of this structure has been undertaken.

The mobilities of 2DHG's in SiGe channels, formed by remote doping, were, initially, limited by Cu contamination. This Cu contamination appears to have been gettered, from the SiGe channel into a B doped Si region, by the use of a growth

interruption. This effect was also thought to be growth related. The Cu contamination was reduced by modifications to the Ge source. Subsequently, a strong dependence of μ(4K) on T_s has been found, which is believed to be associated with the reduction of interface charge at high T_s. Again, this effect is thought to be growth related. The interface charge may be associated with metal contamination. The highest reported μ (4K) in a SiGe channel 2DHG has been measured in a structure grown at T_s=850C, which is considerably higher than is usually used for strained layer growth. This mobility is at least double the highest previously reported. At higher T_s, long range ripple has been observed at the confining Si/SiGe interface. The mobility of the 2DHG has been examined as a function of channel width and of hole density. As a result, it is believed that this long range interface roughness, possibly in conjunction with interface charge, limits the mobility for $T_s \geq$850C.

There are three common features of the results in this thesis.

i) T_s is a highly influential parameter in Si/SiGe MBE growth.

ii) The effects of increasing T_s cannot always be replicated by post growth annealing.

iii Current MBE Si and SiGe growth techniques produce material with trace contamination, particularly metals, which can dominate the electrical behaviour of both 2D and 3D structures.

It is clear that considerable further work is required, regarding material quality, if MBE is ever to become a viable technique for commercial structures. Solid source MBE is currently very useful as a tool for solid state physics studies. However, even for these applications, a greater understanding of growth processes (i.e. impurity incorporation, defect formation) is required. One advantage of a high vacuum technique, like MBE, is that surface study techniques (RHEED, STM etc) can be applied in-situ. This may prove to be important in order to improve our understanding. However, very useful information can be obtained from carrier lifetime measurements,

and even parallel transport studies (which are often the goal of low dimensional growth, rather than a diagnostic tool) and these studies are worth pursuing.

By using low temperature grown oxides, the Zerbst technique can be readily applied to MBE Si as an important measure of material quality. It would be instructive to perform these measurements to examine the τ_g versus T_s dependence as a function of growth rate and, also, to assess the influence of substrate cleaning.

There is currently a trend towards minimising thermal budgets in VLSI processing. Given that MBE Si and SiGe has been found to be relatively stable to post growth annealing, it seems likely that, in the near future, conventional processing will ensure the integrity of high resolution structures. This will be a major advantage for studies of MBE material quality. The recombination lifetime is highly sensitive to electrically active defects, so may have a useful role in assessing the effects of such processing on MBE structures. However, since the requirements of processing high resolution structures and conventional processing techniques may soon converge, this area is not currently a high priority.

Remote doped structures are expected to have many applications in future semiconductor devices. The highest values of low temperature 2DEG mobility in SiGe structures are now comparable with those found in III-V systems. Room temperature 2DHG mobility improvements might be a significant achievement finding application in CMOS technology. It will prove interesting to vary T_s during the growth of SiGe channel structures, in order both to minimise interface charge scattering and to prevent the onset of long range interface roughness. (Stop Press: by continuing this line of study, workers at Warwick have recently obtained $\mu(4K)=1.8\times10^4$ $cm^2V^{-1}s^{-1}$ in a similar structure.)

It is the author's opinion, that the most important challenge for future material studies will be to devise an effective gettering process for high resolution structures. It is intriguing to consider that this might be achieved, in part, by the use of growth

interruptions. The effects of these growth interruptions might be studied as a function of growth schedules and by introducing C, O or H during the interruption. Rapid thermal processing techniques may also prove to be of great value.

References

Abrahams E, Anderson P.W, Licciardello D.C, Ramakrishnan T.V, 1979, Phys Rev Lett 42, 673

Abstreiter G, Brugger H, Wolf T, Jorke H, Herzog H-J, 1985, Phys Rev Lett 54, 2441

Al'tschuler B.L, Khmelnitskii D, Larkin A.I, Lee P.A, 1980, Phys Rev B 22, 5142

Al'tshuler B.L and Aranov A.G, 1985, "Electron-Electron Interactions in Disordered Systems" ed Efros A.L, Pollak M, North Holland, Amsterdam

Ando T, Fowler A.B, Stern F, 1982, Rev Mod Phys 54, 437

Anderson P.W, 1958, Phys Rev 109, 1492

Andrews J.M and Phillips J.C, 1974, Phys Rev Lett 35, 56

Angello S.J, 1958, Proc IRE 46, 968

Bailey W.E, Bowling R.A, Bean K, 1985, J Electrochem Soc 132, 1721

Bardeen J and Brattain W.H., 1948, Phys Rev 74, 230

Bean J.C, 1984, J. Crys Grwth 70, 444

Bean J.C, "Silicon Based Semiconductor Heterostructures" in Si MBE, ed Kasper E and Bean J.C, CRC Press, Florida

Bean J.C, 1984, Jnl Crys Grwth 70, 444

Bean J.C and Sadowski E.A, 1982, J Vac Sci Technol 20, 137

Bean J.C., Becker G.E, Petroff P.M, Seidel T.E, 1977, J Electrochem Soc 48, 907

Becker G and Bean J.C, 1977, Appl Phys Lett 48, 3395

Bergland C.N, 1966, IEEE Trans Elec Dev ED13, 701

Bergmann G, 1984, Phys Reports 107, 1

Birrattella M.S, Neugroschel A, Lindholm F, 1979, IEEE Trans Elec Dev ED 26, 1361

Biswas R, 1992, (Thesis, Warwick University)

Biswas R, **Smith D.W**, Hopkinson M, Houghton R, Parker E.H.C, Whall T.E, 1988, J Electrochem Soc 88-7, 524

Borland J.O, 1985, Met. Res Soc Symp Proc, 36, 269

Brantley W.A, 1973, J Appl Phys 44, 1534

Braun S and Grimeiss H.G, 1972, J. Appl Phys 44, 2789

Brighten J.C, Hawkins I.D, Peaker A.R, Parker E.H.C, Whall T.E, 1993, J.Appl Phys 74 (to be published)

Brighten J.C, 1993, Thesis (Warwick University)

Burton W.K.,Caberra N,Frank F.C. 1951, Philos Trans R. Soc London Ser A 243, 299

Butcher P.N, 1986, " Crystalline Semiconducting Materials and Devices" ed Butcher P.N, March N.H, Tosi M, Plenum (New York)

Byczkowski M and Madigan J.P, 1957, J Appl Phys 28, 878

Calzolari P.U, Graffi S and Morandi C, 1977, Sol State Electron 20, 205

Calzolari P.U and Graffi S, 1972, Sol State Elect 15, 1003

Chattopadhyay D and Quiesser H.J, 1981, Rev of Mod Phys 53, 745

Claeys and Vanhellemont, 1993, Jnl Crys Growth, 126, 41

Correig X, Calderer J, Blasco E, Alcubilla R, 1990, Sol State Electron, 477

Cowley A and Sze S.M, 1965, J Appl Phys 36, 3212

Craven R.A, 1985, Met. Res Soc Symp Proc, 36, 159

Crowell C.R, Shore H.B, LaBate E.E, 1965, J Appl Phys 36, 3843

Delagebeadeuf D and Linh N.T, 1982 IEEE Trans Ed 29, 955

deVisschere, 1986, Sol State Electron 29, 1161

Dingfen Wu, Dening Wang, Heime K, 1986, Sol State Electron 29, 489

Dingle R, Stormer H.L, Gossard A.C, Wiegmann W, 1978, Appl Phys Lett 33, 665

Dohler G, 1986, IEEE J of Quantum Elec 22, 1682

Eichhammer W, Vu-Thuong-Quat, Siffert P, 1989, J Appl Phys 66, 3857

Emeleus C.J, Whall T.E, **Smith D.W,** Kubiak R.A, Parker E.H.C, Kearney M.J, 1992 Thin Solid Films 222, 24

Emeleus CJ, Whall T.E, **Smith D.W,** Kubiak R.A, Parker E.H.C, Kearney M.J, 1993, J Appl Phys 73

Emeleus CJ, Whall T.E, **Smith D.W,** Mattey N.L, Kubiak R.A, Parker E.H.C, Kearney M.J, 1993, Phys Rev B15 47,

Erginsoy C, 1950,Phys Rev 79, 1013

Fahey P.M, Griffin P.B, Plummer J.D, 1989, Rev Mod Phys Vol 61 no 2, 289

Farley C.W, Crook G.E, Kesan, V.P, Block T.R, Stevens H.A, Matford T.J, Neikirk D.P, Streetman B.G, 1987, J. Vac Sci Technol. B5, 1374

Fossum J.G and Lee D.S, 1982, Sol State Electron 25, 741

Foxon C.T, Harris J.J, Hilton D, Hewett J, Roberts C, 1989, Semicon Sci Technol 4, 582

Gershenzon E.M, Mel'nikov A.P, Rabinovich R.I, 1985, "Electron-Electron Interactions in Disordered Systems" ed Efros A.L, Pollak M, North Holland, Amsterdam

Ghez R. and Iyer S.S 1988, IBM J Res Develop 32(6), 804

Givitskas V, Willander M, Noreika D, Petrauskas M, Knall J, Ni W.X, 1988, Semicon Sci Technol 3, 1116

Goetzberger, 1966, Bell System Tech Jnl. 45, 1097

Gold A and Dolgopolov V.T, 1986, Phys Rev B 33, 1076

Gonzalez F and Neugroschel A, 1984, IEEE Trans Elec Dev ED 31, 413

Grove A.S and Fitzgerald A.J, 1966, Sol State Electron 9, 783

Hall R.N and Racette J.H, 1964, J Appl Phys 35, 379

Harris J.J, Pals J.A, Woltjer R, 1989, Rep Prog Phys 52, 1212

Heiman F.P, 1967, IEEE Trans Elec Dev ED14, 781

Hess K, 1979, App Phys Lett 35, 484

Higashi G.S, Bean J.C, Buescher C, Yadvish R, Temkin H, 1990, Appl Phys Lett 58, 2560

Higashi G.S, Bean J.C, Buescher C,Yadvish R, Temkin H, 1990, Appl Phys Lett 58, 2560

Higgs V, Lightowlers E.C, Davies G, 1989, Semicon Sci Technol 4, 593

Higgs V, Davies G, Kubiak R.A.A, 1990, Mat Sci Forum 65-66, 351

Hogarth 1986 Inst Phys Conf "Contacts" London.

Hopkins R.H and Rohatgi, 1986, J Crys Grwth 75, 67

Houghton R, 1991, Thesis, Warwick University

Houghton R.F. Patel G, Leong W.Y, Whall T.E, Parker E.H.C, Kubiak R.A.A, Nayler R, 1987, J Crys Grwth 81, 326

Huang J.S.T, 1970, Proc IEEE 58, 1849

Iyer S.S, Patton G.L, Stork M.C, Meyerson B.S, Harame L, 1989, Appl Phys Lett 36, 2043

Johnson W.C and Panousis P.T, 1971, IEEE Trans Ed ed-18, 965

Jorke H, Kibbel H, Schaffler F, Casel A, Herzog H-J, Kaspar, 1989, J Crys Gwth 95, 484

Kang J.S and Schroder D.K,

Kaspar E, Herzog H-J, Kibbel M, 1975, Appl Phys Lett 8,199

Kaspar E and Bean J.C, 1988, Silicon Molecular Beam Epitaxy, Vol1, CRC Press, Florida

Kawaji S, 1978, Surf Sci 73, 46

Kern W and Rosler R.S, 1977, J Vac Sci Technol. 14, 1082

Khorran S, Chern C.H, Wang K.L, 1991, MRS Proc 220, 181

Kimura A and Lee C.A, 1975, Solid State Electron. 18, 901

Kingston R.H, 1954, Proc IRE 42, 829

Kubiak R.A.A, Leong W.Y, Parker E.H.C, 1985, J. Electrochem Soc 132, 2738

Kubiak R.A.A, Newstead S.M, Leong W.Y, Houghton R, Parker E.H.C, Whall T.E, 1987, Appl Phys A 42, 197

Lang D.V, 1974, J. Appl Phys 45, 3023

Laughlin R.B, 1981, Phys Rev B 23, 5632

Lee P.A and Ramakrishran T.V, 1985, Rev Mod Phys 57, 287

Lescronier D, Paugum J, Richou F, Pelous G, Beniere F,1981, J Appl Phys 51, 1036

Lewis D.C, 1975, Sol State Electron 18, 87

Liu, 1988 (priv comm in Marsh O, 1988, Proc Electrochem Soc 88-8,)

Manasevit H.M, Gergis I.S, Jones A.B, 1982, Appl Phys Lett 41, 464

Manku T and Nathan A, 1991, Phys Rev B 43, 12634

Mattey N.L, Hopkinson M, Houghton R.F, Dowsett M.G, McPhail D.S, Whall T.E, Parker E.H.C, Booker G.R, Whitehurst J, 1990, Thin Sol Film 184, 15

McGill T.C and Baron R, 1975 , Phys Rev B 11, 5208

Mii Y.J, Xie Y.H, Fitzgerald E.A, Monroe D, Thiel F.A, Weir B.E, Feldman L.C, 1991, Appl Phys Lett 59, 1611

Mishima T, Fredriksz C.W, van de Walle G.F.A, Gravesteijn D.J, van den Heuvel R.A, Van Gorkum A.A, 1990, Apl Phys Lett 57, 2567

Miyake M and Harada H, 1981, Jap J Appl Phys 20, L797

Mori S and Ando T, 1980, J. Phys Soc Jpn 48, 865

Mott N, 1978, Sol State Electron 21, 1275

N.L.Mattey 1991, Thesis Warwick University

Nakagawa K and Miyao M, 1991, J. Appl Phys 69, 3058

Nauka N, Logowski J, Gatos H.C, Li C-J, 1985, Appl Phys Lett 46, 673

Ni W-X and Hansen G.V, 1990, Phys Rev B 42, 3030

Nicollian and Brews 1980, MOS , Wiley, New York.

Norton P, Braggins T, Levinstein H, 1973, Phys Rev B 8, 5632

Nutzel J.F, Meier F, Friess E, Abstreiter G, 1992, Thin Solid Films 222, 150.

Orton J.W and Blood P, 1990, "The Electrical Characterisation of Semiconductors: Measurement of Minority Carrier Properties", Academic Press, London

Ota Y, 1983, Thin Sol Films 106, 1

Parker E.H.C.and Whall T.E., 1988, J Electrochem Soc 88-8, 347

Parry C.P, Kubiak R.A.A, Newstead S.M, Whall T.E, Parker E.H.C, 1991, MRS Proc 220, 103

Parry C.P, Kubiak R.A.A, Newstead S.M, Whall T.E, Parker E.H.C, 1992, J App Phys 71, 118

Peaker A.R, Sidebotham E.C, Hamilton B, Pawlick M, 1988, Proc Electrochem Soc 88-8, 392

Pearce C.W. and McMahon R.G, 1977, J. Vac. Sci Technol 14, 40

People R, Bean J.C, Lang D.V, Sergent A.M, Stormer H.L,Wecht K.W, Lynch R.T, Baldwin K, 1984, 45, 1231

People R and Bean J.C, 1986, 48, 538

Pidduck A.J, Robbins D.J, Cullis A.G, Leong W.Y, Pitt A.M, 1993, Thin Solid Films (to appear)

Pindoria G, Kubiak R.A.A, Newstead S, Woodruff D.P, 1990, Surf Sci 234, 17

Pindoria G, 1990, Thesis (Warwick University)

Pindoria G., Houghton R, Hopkinson M, Whall T.E., Kubiak R.A.A, Parker E.H.C. 1990, J. Vac. Sci. Technol. B8, 21

Polignano M.L, Cerofolini G.F, Bender H, Claeys C, 1988, J Appl Phys 64, 869

Powell A.R (a), Kubiak R.A.A, Whall T.E, Parker E.H.C, Bowen D.K, 1991, Mat Res Soc Symp Proc vol 220, 115

Powell A.R (b), Kubiak R.A.A, Whall T.E, Parker E.H.C, Bowen D.K, 1991, Mat Res Soc Symp Proc vol 220, 277

Queisser, 1978, SolState Electron 21, 1495

Ravi K.V, 1981, "Imperfections and Impurities in Semiconductor Silicon", Wiley, New York.

Rees , 1985, Sol State Electron 28, 517

Rozgonyi G.A, Salih A.S.M, Radzimski Z, Kola R.R, Honeycutt, Bean K, Lindberg K, 1987, J Crys Grwth 85, 300

Sakaki H, Node T, Hirakawa K, Tanaka M, Matsusue T, 1987, Appl Phys Lett 51, 1934

Salih A.S, Kim H.J, Davis R.F, Rozgonyi G.A, 1985, Appl Phys Lett 46, 419

Schaffler F, Tobben D, Herzog H-J, Abstreiter G, Hollander B, 1992, Semicon Sci Technol 7, 260

Schimmel D.G, 1979, J. Electrochem Soc 126, 479

Schlegel F, 1967, IEEE Trans Ed ed-14 no.11, 728

Schockley W and Read W.T (Jnr.), 1952, Phys Rev 87, 835 and Hall R.N, 1952, Phys Rev 86, 600

Schroder D.K, 1984, Sol State Electron 27, 247

Schroder D.K and Nathanson H.C, 1970, Sol State Electron 13, 577

Schubert E.F, Fischer A, Ploog K, 1986, IEEE Trans Ed 33, 625

Schubert E.F and Ploog K, 1985, Jpn J Appl Phys 24, L608

Schuberth G, Schaffler F, Besson M, Abstreiter G, Cornik E, 1991, Appl Phys Lett 59, 3318

Schwettman F.N, 1974, J. Appl Phys 54, 1918

Sclar N, 1956, Phys Rev B 104, 1559

Sebesteyen T, 1982, Sol State Electron 25, 543

Shannon J., 1979, Appl. Phys. Lett 35, 63

Shockley W, 1949, Bell Syst. Tech J. 28, 435

Sidebotham E.C., Peaker A.R., Hamilton B.,Hopkinson M, Houghton R, Patel G, Whall T.E, Parker E.H.C. 1988, J Electrochem Soc 88-8, 360

Smith D.W, Hopkinson M, Houghton R.F, Whall T.E, Parker E.H.C, 1988, J Electrochem Soc 88-7, 538

Smith D.W, Houghton R.F, Parker E.H.C, Whall T.E, 1990, Thin Sol Films 184, 177

Sollner T.C.L.G, Goodhue W.D, Tannenwald P.E, Parker C.D, Peck D.D, 1984, Appl. Phys Lett 43, 588

Sparks D.R, 1987, J Electrochem Soc 134 , 458

Spirito P and Cocorullo G, 1985, IEEE Trans Elec Dev ED 32, 1708

Stacy W.T., Allison D.F, Wu T.C, 1981, Semiconconductor Science ed by Huff and Kriegler , Electrochem. Soc,. 334

Stolwijk N.A, Frank W, Holzl J, Pearton S.J, Haller E.E, 1984, J Appl Phys 57, 5211

Tang Y.S, Wilkinson C.D.W, Sotomayor C.M, **Smith D.W,** Whall T.E, Parker E.H.C, 1992, Superlatt. Microstruc. 12, 535

Tanner B.K and Bowen K, 1993, J Crys Gwth 126, 1

Taylor S, Zhang J, Ecclestone W, 1993, Semicon Sci Technol 8 , 1426

Tsui D.C, Stormer H.L, Gossard A.C., 1982, Phys Rev Lett 48, 1559

Van de Walle C.G and Martin R.M, 1986, Phys Rev B 34, 5621

Van Gorkum A, 1992, priv comm

Van Gorkum A.A, van de Walle G.F.A, vand den Heuvel R.A, Gravesteijn D.J, Fredriksz , 1990, Thin Sol Films 184, 207

Van Wees B.J, van Houten H, Beenakker C.W.J, Williamson J.G, Kouwenhoven L.P, van der Marel D, Foxon C.T, 1988, Phys Rev Lett 60, 848

Venkatarmaran,V, Liu C.W, Sturm J.C, 1993, Elec. Mat. Conf, Santa Barbara

Vick G.L and Whittle K.M, 1969, J. Electrochem Soc 116, 1142

VLSI Technology by Sze

Von Klitzing K, Dorda G, Pepper M, 1980, Phys Rev Lett 45, 494

Wei C-Y and Wooodbury H.H, 1985, IEEE Trans Elec Dev ED 32, 957

Wharam D.A, Thornton D.J, Newbury, R, Pepper M, Ahmed H, Frost J.E.F, Hasko D.G, Peacock D.C, Ritchie D.A, Jones G.A.C, 1988, J Phys C 21, L209

Whelan, 1969, Sol State Electron 12, 963

Wieder H.H, 1979, Laboratory Notes on Electrical and Galvanomagnetic Measurements", Elsevier, Oxford

Yau and Sah, 1974, Sol State Electron 17, 193

Yu A.Y.C, 1970, Sol State Electron 13, 239

Zeindl H.P, Wegehaupt T, Eisele I, Oppolozer H, Reisinger H, Tempel G, Koch F, 1987, Appl Phys Lett 50, 1164

Zerbst M,1966, Z. Angew Phys 22, 30

Ziegler K and Klausman E, 1974, Appl Phys Lett 26, 400